DOCUMENT RESUME

ED 136 596 FL 008 469

AUTHOR	Wanat, Stanley F., Ed.
TITLE	Issues in Evaluating Reading. Linguistics and Reading Series: 1. Papers in Applied Linguistics.
INSTITUTION	Center for Applied Linguistics, Arlington, Va.
PUB DATE	Mar 77
NOTE	77p.
AVAILABLE FROM	Center for Applied Linguistics, 1611 N. Kent St., Arlington, Virginia 22209 ($4.95)
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS	*Applied Linguistics; Cultural Factors; Intelligence; Intelligence Quotient; Language Attitudes; Language Research; *Linguistics; Linguistic Theory; *Reading; Reading Readiness; Reading Readiness Tests; *Reading Research; Reading Skills; *Reading Tests; Teacher Attitudes

ABSTRACT

Six papers are included in this collection. "How Misconceptions about Language Affect Judgments about Intelligence," by Roger W. Shuy, and "The Cultural Context of Learning to Read," by R. P. McDermott, deal with contexts in which evaluation of a student's reading are made: a teacher's direct evaluation of a student's reading performance, and a teacher's interpretation of a student's performance on a reading test take place within the context of what the teacher thinks of that student's general level of cognitive abilities. "Criteria for Evaluating Readiness: 'When Is a Child Intelligent Enough to Read?'," by Stanley F. Wanat, and "Consumer Awareness in Testing Reading," by William Eller and Roger Farr, discuss the content, form, and use of standardized reading tests. In his paper entitled "What Should Early Reading Tests Measure?", J. Jaap Tuinman stresses the need for close correspondence between what is tested and what is taught, and "IQ Is and Is Not Related to Reading," by Harry Singer, deals with the variable relationship between IQ and reading dependent on the nature and difficulty of the task, the capabilities of the reader, the time allowed for learning, the quality of instruction, and the nature of the tests used for assessing intelligence and reading. Eight pages of references conclude the volume. (CLK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

Papers in Applied Linguistics

ED136596

LINGUISTICS AND READING SERIES: 1

Issues in Evaluating Reading

edited by Stanley F. Wanat

Roger W. Shuy, General Series Editor

FL008469

U.S. DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CENTER FOR APPLIED LINGUISTICS

2

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Papers in Applied Linguistics

LINGUISTICS AND READING
SERIES: 1

Issues in Evaluating Reading
edited by Stanley F. Wanat

Roger W. Shuy, General Series Editor

3

Center for Applied Linguistics

The editors wish to express their thanks to all the authors who have contributed to this volume. They also wish to express their special appreciation and thanks to Diane Bartosh of the CAL publications program for her careful and painstaking help in the copy-editing of the volume and in its final preparation for publication.

Library of Congress Catalog Card No.: 77-75685
ISBN: 87281-060-7

April 1977
Copyright © 1977
By the Center for Applied Linguistics
1611 North Kent Street
Arlington, Virginia 22209

Printed in the U.S.A.

Preface

The important relationship of linguistics to reading has long been recognized but seldom has it been adequately exploited. Many linguists feel that when people read, they call on what they know about language form and use in ways which we, as yet, little understand. Many reading specialists, likewise, feel that reading is some type of language processing operation in which the language takes the form of print on the page. Although both the fields of linguistics and reading are concerned with meaning, there has been little sharing of ideas across the fields. A great deal of mutual focus, in the past, has been on letter-sound correspondence. This focus has led many teachers to consider the contribution of linguistics to be little more than phonology. Some reading people, in fact, erroneously think of this contribution as "the linguistic approach."

As the title indicates, this series is dedicated to the bringing together of linguistics and reading. It is intended to present both technical and practical matters which contribute to the comprehension of the printed page. The various issues in the series will be theme oriented, centering around a specific concept or approach for which both the fields of linguistics and reading can benefit. The ultimate goal of the series is to open insights about reading to linguists and insights about linguistics to reading specialists.

Roger W. Shuy
March 1977

Evaluating Reading: An Introduction

Are current procedures for evaluating reading adequate? Or do the reading tests so widely used in the schools provide inadequate, useless, and misleading information? Is there a need for developing more adequate criteria for evaluating reading? Many opinions and impressions have been voiced in answer to these questions. The papers in this collection provide analysis and evidence for their further exploration. This introduction discusses some major characteristics of current procedures for evaluating reading, and then notes the contribution that each of the papers in this collection makes to the development of criteria for evaluating reading.

What are some of the characteristics of commercial tests widely used in the schools? A test deals with just a small segment of a person's behavior, and a test score can provide only an estimate of a person's performance. Furthermore, "teachers talk sometimes of 'second grade' level or 'fourth grade' level as if these labels represented well-defined and widely accepted standards. They do not (Farr, 1971)." Also, the test may not truly sample relevant real-world reading behaviors--or the other behaviors the test claims to evaluate.

Following is a listing of major limitations of standardized tests: "(a) The very fact that a test is 'standardized' in terms of administration and scoring may make it inappropriate for use with certain groups or individuals. (b) The test-maker's quest for brevity, which unfortunately, but pragmatically enhances the salability of tests in some circles, may result in unrealistic time limits and a choice between depth and breadth in sampling. (c) Group administration may work to the disadvantage of certain individuals...a child who fails to understand one or two words in a set of directions may be unable to respond to any of the items, which he may or may not have known. (d) The format of the test may restrict the type of items used....Certain behaviors are not adequately sampled with multiple-choice items (Otto, 1973)."

Following is a critique of a group of tests designed to analyze different components of reading skill: "Davis [see Davis, 1944] has carried out some of the most meticuluous research on the differentiability of different types of reading skills....His tests had to be quite short, since he was measuring some 8 different aspects of reading, so that the reliabilities of these tests were only .58 and .59. The correlation between them had an average value of .45 in several sets of data. Given these values, the betting odds are only 1 to 4 that a differences of one standard deviation between scores on the two tests is 'real.' [The test reviewer provides the following definition with regard to such a difference.]...One full standard deviation of difference is a difference that would correspond roughly to falling at the 70th percentile of a group on one measure and the 30th on the other (Thorndike, 1973)." This is a serious criticism of "some of the most meticulous research" in the area. When one finds a 40 percentile difference between two scores, the odds are only 1 to 4 that the difference between these two scores is real.

On the basis of Thorndike's (1973) analysis of another test, the Stanford Diagnostic Reading Tests, one can estimate that for 20 percent of the students who take the test, the odds are about even that a difference between two scores

is a real difference, and not due to measurement error. For the other 80 percent of the students, there is not even a 50-50 chance that the differences in their test scores are real differences. For about 40 percent of the students taking this test, the odds are 6 to 1 that the "difference" shown by the test is due to measurement error.

Another reviewer provides the following critique of the Clymer-Barrett Prereading Battery (CBPB) Form A: *"The authors do not seem to know what they would like to do...the authors do not know what they have here.* Sometimes this is a reading readiness test; sometimes it is not. They open their manual on page 5 with their definition of a reading readiness test; it 'provides information about a child's preparedness for learning to read.' They continue that their test is intended to be consistent with their 'beliefs about the nature of reading readiness.' Then on page 28 they turn around and say proudly that the data show that CBPB is *different* from various (unspecified) 'older, recognized tests of the same general kind.' Correlational studies with these other readiness tests and CBPB apparently run in the range .55 to .80; although higher than much of the other correlational data in the manual, [these correlations] are hailed as confirming the idea that CBPB is not just another readiness test....*The authors are most misleading when they claim diagnostic advantages.* The authors state that diagnostic information 'can be secured by analyzing the subtest scores and comparing them with one another'.... The authors do that, and find on page 28 that the subtest intercorrelations are too low to use. The median score is a stunning .24. They then lump together the data from pairs of subtests to raise the subtest intercorrelations. The resulting 'diagnosis' seems fairly worthless. It tells the teacher for instance that the child is needing practice in completing geometric shapes plus the copying of words; these are two very different activities. In the same way, another child may apparently need practice at picking out printed letters in response to spoken letter names plus practice at picking out printed words that match other printed words. Again these two quite different skills are lumped together in the diagnosis. It is rather as if the X-ray could not tell the surgeon whether the broken limb he had to set was an arm or a leg (N. Mackworth, 1974A)."

Another reviewer, in a critique of the same test, raises these cautions: The authors "assert, with no references, that 'there is enough evidence available to suggest that instruction based on differential analysis of the three CBPB part scores will prove more worthwhile than will ignoring these scores and teaching all pupils as though all their skills were equally developed [CBPB manual, page 25].' However, they then list a series of qualifications that seem to render this information useless...(Sticht, 1974)." The reviewer also cautions that there is no clear evidence that two-thirds of the whole CBPB adds much predictive power to that obtained with just the other one-third of the test.

Detailed analysis of four widely used reading tests indicates the following major weaknesses (see Wanat, 1974, 1975): *1. Tests fail to include important skills. 2. Tests assess skills irrelevant to reading by including tasks not related to reading. 3. Test instructions telling the child what to do are not clear.*

4. The test may confuse intelligence with reading if, as in the case of one test, the scores obtained with that test correlate as highly with the scores obtained from a set of intelligence tests as they do with the scores obtained from a set of reading achievement tests. The manual accompanying one test notes that four intelligence tests have correlations with this test's total score falling in the same range as correlations between this test's total score and three reading achievement tests. This creates a problem, for this particular test promises to provide a measure of the child's intelligence instead of, or inseparably intertwined with, an assessment of reading (Sticht, 1974; Wanat, 1974, 1975).

5. The test may confuse correlation with causation if, as in the case of many tests, it is implied that a measure which is predictive of (correlated with) later success in reading provides, by virtue of this correlation, an identification of remediable causes of reading difficulties. One test measures the child's recognition of letters of the alphabet. It has often been pointed out that this ability is highly predictive of (correlated with) later success in reading acquisition (Calfee and Venezky, 1972; Lowell, 1970). While a child who is better able to identify the letters of the alphabet is more likely to do well later in reading, there is no evidence that teaching the names of the letters to a child who does less well on this measure will cause the child to be a better reader. By failing to obviously, directly, and non-technically communicate to test consumers the limitations of their testing instruments, test authors provide much of the basis for the misinterpretation and misapplication of test results. While test authors "protect" themselves by noting limitations in their test manuals, those limitations are embedded in a mass of technical language that most test consumers find difficult to evaluate. Unfortunately, this "protection" extends only to the test's authors and publishers, and not to the students who use these tests.

6. The test may have an irrational scoring procedure. For example, in a copy-a-sentence subtest in one reading readiness test, the child loses four out of the seven possible points by failing to skip a line, even though all of the words are copied perfectly. There is no rational basis for having the child lose over 50 percent of the possible points for such a "mistake." No case has been made for relating the ability to skip lines to reading readiness, and it is hard to imagine that teaching a child to skip lines when he is copying sentences will improve the child's reading (Sticht, 1974; Wanat, 1974, 1975).

7. The test may be an inefficient way of gathering information since some tests require considerable time. In estimating the benefits of using a particular test, one must weigh the value of the information gained against the effort required to obtain that information. One must consider the time required to use the test, including the time to select the test, study its manual, administer the test, interpret the results, and figure out their implications for instruction.

8. Items in subtests measure such diverse skills that it is hard to ascertain what is actually being assessed. 9. Subtests have too few items to give reliable information.

10. Subtest scores may give information that has no instructional value if, as in the case of many tests, the items within a subtest are too varied, or if there are not enough items, thus leading to a situation in which differences between subtests scores cannot be used diagnostically (that is, for identifying specific areas of weakness). In one such test, the publisher's manual warns that significance cannot be attached to the separate scores from the subtests. Yet, this test is set up so as to yield subtest scores. Dividing the test into subtests creates the false impression that useful information can be gathered on the component skills. This situation highlights the diagnostic inadequacy of current reading tests. The compilers of one set of test reviews point out that "all of the reading tests reviewed in this book measure global reading or readiness skills. In spite of the titles, the tests are of little diagnostic value (Farr and Anastasiow, 1969)." Another reviewer makes this comment about available diagnostic reading tests: "The confidence [that one can logically place in an observed difference between two scores] is often distressingly low.... If salvation exists, it lies in the fact that most of the actions following from diagnostic judgments are reversible (Thorndike, 1973)." One can, of course, teach reading, or anything else, without first diagnosing specific weakenesses in the learner. Such diagnosis, however, enables one to match instruction to the specific needs of the learner. Successful school programs are characterized by regular diagnostic testing (Clark, 1972), which presumably aids this kind of matching. With respect to the diagnostic inadequacy of reading readiness tests,

it has been noted that "performance on readiness subtests has been put forward as a source of diagnostic information, yet there is no clear evidence of the validity of these measures as diagnostic indicators; nor is it apparent what remedial action should be taken when a child performs poorly on a readiness subtest (Calfee and Venezky, 1972)."

11. The test may provide inadequate instructional guidance by failing to relate test outcomes to specific instructional procedures. An *EPIE Educational Product Report* (EPIE, 1974) comments on one reading test: "This is an excellent tool for diagnosing the verbal reading of students in grades one to seven. What teachers do with the data once they obtain it is a significant but unadressed question." It is difficult to see the usefulness in "an excellent tool for diagnosing" that does not directly lead to specific instructional decisions. One test, for example, notes that the teacher must figure out what to do if the test yields results different from the teacher's own evaluation of the student's performance. If the teacher must figure out what course of action to follow when a test's assessment of a student's performance differs from the teacher's own assessment, then it is difficult to see what the teacher has gained by using this test (Sticht, 1974; Wanat, 1974, 1975).

12. Subtests making up the tests do not measure what they claim to measure. The test author's mislabeling of what the items in a test (or subtest) measure provides the basis for the teacher's mislabeling, misdiagnosis, and consequent mistreatment of the students and their individual reading difficulties. Examination of the contents of tests will reveal the disparity between what the test says it measures, and what the items actually require the students to do.

For example, examination of reading tests in one widely used series of achievement tests will show that the authors claim that these tests provide a measure of the students' reading comprehension. Further examintion of the contents of the reading comprehension section shows six blocks of text-plus questions. One of these blocks consists of items testing the students' ability to use a table of contents. A second block of items tests the students' ability to use an index. Three of the blocks are prose passages, each passage accompanied by a set of questions. Some of the questions with each passage require the students to locate details in the passage. Other questions are more global in nature, requiring the students to make "inference-type" responses. The rationale for the preponderance of one of these question types over the other is not apparent in the test. (Use of the Fry Readability Graph shows a very wide range in the reading difficulty levels of the passages in this particular test.) The last block of text-plus questions involves a text different in kind from the texts in the three preceding blocks, since it includes a lot of numbers. Thus, whether or not one accepts the claim that such a test assesses how well the student comprehends what he reads depends in large measure on (a) one's willingness to accept this hodgepodge of tasks as representative of the real world of reading comprehension and (b) one's ability to feel comfortable with the test's failure to present either a theoretical or empirical rationale for what is included and excluded, and for the varying amounts of emphasis given to the different types of reading tasks.

The following situation should bring into focus another set of circumstances that might cause a test user to feel uncomfortable with this test's claim of measuring reading comprehension: Assume that two students got the same total number of items correct when taking this test. All of the wrong answers of one of the students were on items requiring the student to look for details in the passage, while all of the wrong answers of the other student were on items requiring "inferencing-type" skills. Would one feel comfortable with the claim that the reading comprehension of skills of both students are the same? Since this test gave both students the same score, the test indicates that both students should be treated alike with respect to the instruction to be formulated for them. The selection of tasks to be included in a reading test should be based upon some explicit, coherent, and consistent description of the reading

skills being tested. Tests such as the one discussed above make almost no contribution to guiding instruction. The teacher would obtain more instructionally-relevant information by observing the students' reading behaviors in a more direct fashion.

The twelve weaknesses noted above deal with the *content and form* of a sample of reading tests. Four major problem areas associated with the *use* of currently available reading tests are: *(a) The test instrument itself may be poor,* if it contains deficiencies such as those noted above. *(b) The procedures for administering the test may be inappropriate,* if the test has been designed for use with children who speak only "school English" at home and in school, and the test is administered to children who use a different variety of the language; then, the test results may provide undependable information. *(c) The interpretation of the test results may be inappropriate,* if the test's interpretive scales were developed using an inappropriate "norming" group. "Norm-referenced scales depend upon a specified population of examinees for their meaning...they place an individual on a scale developed to be effective at giving relative position within a population (Woodson, 1973)." Thus, norm-referenced interpretive scales are used to express the test performance of an individual in terms of the performance of a reference (or norming) group. National norms make it possible to estimate how well a given student in, for example, the first grade, performed in comparison to all first grade children in a particular country. Thus, the norm (or "reference") group is all-important, since it gives meaning to the behaviors recorded in the test. But if a test has been normed on a population of upper-middle-class suburban children with a particular linguistic-cultural background, then the scores based on this reference group may be inappropriate if one is trying to evaluate the performance of poverty-level urban (or rural) children with a different linguistic-cultural background. *(d) The educational application of the test results may be inappropriate.* That is, even if the test instrument, the administration procedures, and the interpretation of the test results are appropriate, the testing procedure may still be useless and even damaging to the child if there is an inappropriate application of these results. This would be the case if, for example, the instructional groupings made on the basis of the test results are not subject to frequent review, or if it turns out that a decision is made to not deliver instruction to the low-scoring students.

What do the papers in this collection have to say about the content, form, and use of reading tests, and about the development of criteria for evaluating reading? The papers in this collection stress two themes. *First,* current approaches to evaluating reading are seriously deficient and must be improved. Concrete strategies for improvement are presented in this collection. *Second,* current procedures--informal as well as formal--for evaluating reading display basic misconceptions about language. To improve the teacher's informal (and sometimes unconscious) evaluations, and the test maker's formal evaluations, greater awareness about language is necessary. The deficiencies in current approaches for evaluating reading are basic and pervasive. The intent of this collection is to provide an information base that teachers and researchers concerned with reading and language can use in developing criteria for evaluating reading.

The first two papers in this collection deal with *contexts* in which evaluations of a student's reading are made. A teacher's direct evaluation of a student's reading performance, and a teacher's interpretation of a student's performance on a reading test take place within the context of what the teacher thinks of that student's general level of cognitive abilities (or intelligence). If, for example, an instructor has two students, both of whom have done poorly on some (formal or informal) reading evaluation, and if the instructor believes that one of the students has the intelligence to learn, but the other does not, then it is likely that far more attention and instruction will be given to the first student than to the second. Thus, the teacher's evaluation of the student's

intelligence provides a basis for evaluating the student's reading performance. Shuy's paper discusses how this context for evaluation is distorted by misconceptions about language.

McDermott's paper discusses another context in which evaluations of a student's reading take place. He concentrates on the interactional context in which reading instruction takes place. Patterns of interaction in the classroom may create a situation in which the teacher is unable to make sense of the students' behavior, and the students are unable to make sense of the teacher's behavior. Patterns of interaction in the classroom can determine whether or not an instructor will be able to, or will want to, deliver instruction to a particular student. Similarly, such patterns of classroom interaction can determine whether or not a student will be able to, or will want to, learn from a particular instructor. Thus, the teacher's judgment about a student's intelligence, and the interactional patterns of the classroom are contexts in which reading is evaluated. Both of these contexts for evaluation can be influenced by misconceptions about language. Thus, misconceptions about language can lead one to misjudge a person's intelligence, and, in the case of the classroom, these misjudgments about intelligence can lead to patterns of teacher-student interaction that prevent effective evaluation, instruction, and learning.

The next two papers discuss the content, form, and use of standardized reading tests. Wanat's paper examines factors assessed by widely used reading readiness tests, and readiness factors identified by research on reading and on prerequisites for reading. As a concrete strategy for improving current procedures for evaluating reading, a set of criteria for the design and use of readiness evaluations is developed, and a set of linguistic, cognitive, perceptual, and interpersonal factors relevant to the evaluation of reading readiness is identified.

Eller and Farr's paper provides two reviews of tests for evaluating reading readiness. These reviews are presented in distinctly different styles so as to provide examples of two approaches to developing criteria for evaluating reading tests. The central concern of their paper is to develop consumer awareness in evaluating reading, and they provide a concrete strategy for improving current reading evaluation procedures through their development of an instructional-decision-making approach to evaluating reading. They also develop criteria that test consumers should apply when using the consumer information provided by formal test reviews.

The Wanat and Eller and Farr papers continue Shuy's and McDermott's emphasis on the contexts in which reading evaluation and instruction take place.

Tuinman, in the next paper, stresses the need for close correspondence between what is tested and what is taught--a criterion also noted in the two preceding papers. He, like Eller and Farr, stresses the instructional utility of the information provided by a test as a criterion for evaluating that test. A concrete strategy for improving procedures for evaluating reading put forth is that teachers select, develop, and have confidence in their own informal "real-life" probes (in terms of the everyday learning tasks in that teacher's classroom) for evaluating reading. Such teacher-developed readiness and achievement probes could simply consist of a sample of the tasks that students encounter in her classroom. Tuinman argues that the development of reading tests for the early stages of reading should be based on an explicit description of reading acquisition. Thus, a criterion for evaluating tests of reading in its early stages should be the adequacy of the description of reading acquisition that underlies each test.

The final paper in this collection deals with relations between IQ and reading. The topic of intelligence was also addressed directly by Shuy and indirectly by McDermott in their discussions of contexts in which evaluations of reading are made. The topic of intelligence was also addressed by Wanat and Tuinman in their discussions of approaches to readiness assessment. The use of measures of intelligence to predict reading performance is well-documented in the

literature. Shuy has pointed out how evaluations of language lead to judgments about the intelligence of that speaker. Our (mis)perceptions of an individual's language, intelligence, or reading often lead us to (mis)judgments about his performance or potential with respect to one or both of the remaining factors. References to relations among language, intelligence, and reading, and references to the roles that actual or assumed relations among them play in the design, use, and interpretation of student evaluation measures, occur frequently in this collection.

In the last paper, Singer identifies reasons behind the belief that IQ is related to learning to read. He discusses the factors that determine the extent of relationship between IQ and reading, and he explains the conditions under which the relationship between IQ and reading is maximized, and the conditions under which this relationship is minimized. He provides a concrete strategy for improving current procedures for evaluating reading by explaining how absolute as well as relative measures of student learning can be obtained, without the need for schools to turn to tests other than those they currently use. Singer's "IQ Is And Is not Related to Reading" is the first recipient of the International Reading Associations' Albert J. Harris Award for outstanding contribution to diagnosis in reading.

Stanley F. Wanat
March 1977

Contents

Preface.....iii

Evaluating Reading: An Introduction.....v

How Misconceptions about Language Affect Judgments about Intelligence.....1
Roger W. Shuy

The Cultural Context of Learning to Read.....10
R.P. McDermott

Criteria for Evaluating Readiness: "When Is a Child Intelligent Enough to Read?".....19
Stanley F. Wanat

Consumer Awareness in Testing Reading.....28
William Eller and Roger Farr

What Should Early Reading Tests Measure?.....41
J. Jaap Tuinman

IQ Is and Is not Related to Reading.....43
Harry Singer

References.....56

How Misconceptions about Language Affect Judgments about Intelligence

Roger W. Shuy,
CAL & Georgetown University

Despite the great strides being made in our knowledge of the tremendous range of variation that exists within our native American English language, we still know relatively little about subjective reactions to that variation. This paper presents some evidence that subjective evaluations about language use can be whimsical, if not dangerous, and that the schools should tool up to their own potential for misassessing children's intelligence and potential on the basis of such reactions.

As is often the case in the U.S., our strength has led us to a new weakness. That is, our concept of practically universal education has brought to our schools people whose culture and language is in many ways different from that of the presumed school norm. Thus, while it is morally good to educate freely and democratically, the admixture of children resulting from this democracy has brought with it an admixture of language variation and, humans being what they are, a plethora of accompanying prejudices.

Recently in a teacher in-service workshop, I played a tape recording of a working-class man to a group of teachers. I then asked the participants what they could tell about that man on the basis of listening to the recording. Not surprisingly, one teacher said he could tell from the recorded speech sample that the speaker was illiterate and of low intelligence. Needless to say, the teacher was not able to provide evidence from the tape to clearly indicate illiteracy or stupidity. His reaction, instead, was based on a presumed correlation of the speaker's nonstandard English with these characteristics, however insupportable the data may have been.

Of course, such subjective reactions are not limited to American teachers. The Cakchiquel Indians of Guatemala speak a language that is so expressive in its verbal system that a single verb can be conjugated in over 100,000 forms. Yet, until very recently, Cakchiquels were considered inferior and tongue-tied by the Guatemalan Spanish-speaking population.

Since it is sometimes easier to see our own situation through conditions which are not quite so close to us, let us examine the Cakchiquel case and try to determine why speakers of such a complex and well-developed language could be considered ignorant for using it.

Clearly, the conditions are totally non-linguistic. The local Indians were considered "beasts of burden" and, at best, semi-human. Laws which permitted a form of serfdom based on indebtedness had bound most of them to a plantation existence and a life of illiteracy. Few Cakchiquels could read or write and, in any case, the only books were in Spanish. The resulting ignorance left them under the influence of witchcraft and superstition. The ruling Spanish class would have nothing to do with such people and would not even go so far as to shake hands with them. It was not until very recently that the Cakchiquel language was studied and analyzed by outsiders and discovered to be the marvelous, complex, and expressive tongue that it is. If it was difficult for the Cakchiquels to learn Spanish, it was at least equally difficult for the Spanish speakers to learn Cakchiquel, a situation which had been totally overlooked by Guatemalans for centuries. Tongue-tied indeed!

One basic misconception about language which can affect judgments about intelligence, then, is that a language or dialect used by a minority reflects some inability to use language effectively or to provide evidence of underlying intelligence. At best, such a judgment can be considered correlational.

Hearing a Cakchiquel speak Cakchiquel would imply serfdom, drunkenness, witchcraft, and social isolation. But it in no way would imply lack of intelligence on the part of its speakers. Nor would it justify accusations of primitiveness against the language. Most absurd of all, however, would be to imply that speakers of Cakchiquel are linguistically deprived, especially when their language has been subsequently shown to have such beauty and expressiveness. What is clearly involved here is a classic case of ethnocentrism and prejudice.

One might ask whether such a remote case as this has any bearing on how we make judgments about intelligence through language in the U.S. today. The first question to be asked is whether or not judgments about intelligence are being made on the basis of language alone. This is difficult to assess for two reasons. For one, language seldom, if ever, exists in a situation whereby it is not accompanied by other kinds of information. Second, it is difficult to determine whether or not intelligence is the feature which is being judged.

One attempt to get at the answer to the question of how people use language to make judgments about other people was made in 1969 by Shuy, Baratz, and Wolfram. In this study, Washington, D.C. employers were asked to listen to representative tape-recorded speech of local teen-aged and adult males from the complete range of socio-economic status groups in the city. In general, the employers reacted quite predictably to the taped speech samples. They said that they would hire the more standard speakers and that they would be wary of hiring the lower-status speakers. Although most of the employers denied that speech had anything to do with hiring procedures, their assignment of potential jobs to various speakers paralleled exactly the categories of socio-economic status which these speakers represented. The better jobs invariably went to the standard English speakers. Those who were judged unemployable were invariably those with a lesser degree of standard English.

Although the preceding example did not specifically relate misconceptions about language to judgments about intelligence, it hinted strongly at a relationship between language and human potential, and it suggested a kind of informal follow-up study which came closer to questions involving perceptions about intellect as a reflection of language use. One employer, for example, observed that speech tells "...an unmistakable story about a person's... mental alertness." Such an observation has a familiar ring to it for those of us who have been working in teacher education. One of the greatest obstacles in educating teachers about how language works are the various myths about language that continue to exist. One such myth has been developed and perpetuated by a group of educational psychologists who drew upon the early (and rather unclear) writings of the British social psychologist, Basil Bernstein, and concluded that middle-class language is more abstract, more flexible, more detailed and more subtle than working-class speech, and that the lower-classes are more concerned with the here-and-now. Such evidence, now refuted by Bernstein, led to the development of practical programs of many sorts. Particularly prominent among such programs is the one developed by Siegfried Engelmann and his associates, based on empirical research with Carl Bereiter among four-year-old Black children in Illinois. This research suggested that the Black children could not ask questions or make statements of any kind. They were said to communicate primarily through gestures, single words in isolation, and in syntax which was at best badly connected. The classic sentence for linguists to quote has been Bereiter's (1966) observation: "...the language of culturally deprived children is not merely an underdeveloped version of standard English, but a basically non-logical mode of expressive behavior." Teachers were urged to proceed with such children as though they had no language at all, and to train the children to speak in fully explicit, formal language.

The reaction to such research and development has been utterly astounding. The myth attracted immediate followers. Linguists have known for years that language is little understood by the general public, but we had no idea that social interaction was so poorly perceived. Anyone who has ever done fieldwork in child language can attest to the ease with which children can be intimidated into silence or "nonverbality." A child learns very early that the name of the game is to be right as often as possible and wrong as seldom as possible. One way to avoid being wrong is to keep one's mouth shut. As Labov (1972) has clearly shown, there is no reason to doubt Bernstein's data: "On the contrary, they give us a very clear view of the behavior of student and teacher which can be duplicated in any classroom." Whenever a child is in an asymmetrical relationship with an adult, the possibility of accurately measuring his verbality is considerably reduced. Yet such measures of his verbality are constantly being made and used as an explanation for his poor showing in school. The child's poor performance on such measures is excused as a result of squalor, noise, excessive television-watching, inadequate genes or what-have-you. Linguists who have attempted to obtain representative speech samples from children of any race or socio-economic status have developed fieldwork strategies which have yielded amazingly contradictory evidence to that obtained by Engelmann and Bereiter. Simple devices to relax the child include talking to him with his friends present, sitting on the floor instead of at a desk or in a chair, and adding a supply of popcorn and cokes.

My own search for the "nonverbal" child has taken me into many American classrooms (although, in fact, exploratory data from Australia revealed an almost identical situation there). To demonstrate how pervasive the myth of non verbality is, permit me to describe a typical visit to a typical, American elementary school which is attended by some minority children.

RS: I hear that there are a lot of nonverbal children in the schools. Do you have any here?

Teacher (hesitatingly): Yes.

RS: Gee, I'd like to see one. Could you point out this child to me?

Teacher (Pausing): Well..., I think John back there would be a good example.

The point of the above dialogue is simply that the teacher had heard that nonverbal children exist. Since they exist, the teacher felt that she must have one. If she had one, it had to be the one who is her greatest problem. This latter conclusion leads teachers to nominate a wild variety of school-problem cases. Most are minority children. Some are discipline problems, though they indicate many signs of intelligence. Some are quiet and withdrawn. Ninety percent are males. Some give clues to a possible low intelligence. But *none*, absolutely none, gives evidence of nonverbality in any sense in which the term might be legitimately used. My typical next step is to ask to talk privately with the candidate somewhere, either alone or with a friend of his. Such a maneuver almost invariably yields a 10 or 15 minute sample of continuous discourse that easily refutes any accusation of nonverbality. If I have tape recorded such sessions, I generally play the tape to the teacher. What comes out of such an exchange is hard to evaluate, but it is relatively certain that the teacher will at least hesitate before assigning labels such as nonverbal so easily again.

If potential employees can be ranked for possible employment into accurate socio-economic groupings using only the clues of language, it follows that

teachers are quite likely to do similar rankings. Little is known about the subjective evaluations of teachers related to the language of their students, but the few bits of evidence currently available point to a morass of confusion. A study done by Anne Hughes (1967) asked 30 urban teachers to discuss the language problems of minority children as revealed by tape recordings. Most of the teachers complained of the students' limited vocabulary, occasionally commenting that it related to their limited experiences. Many teachers lamented the lack of "complete thoughts" produced by their students. It is not the point of this paper to catalogue the lack of linguistic sophistication evidenced by the teachers in this study, but rather to point out that subjective judgments about children are being made using language as a primary guide.

Likewise, in an informal study done through the combined efforts of the Center for Applied Linguistics and the University of Nebraska a few years ago, elementary school reading teachers from Boston, Dallas, Atlanta, Washington, D.C., and rural, central Ohio were brought together for a three-week workshop. The first task of these teachers was to evaluate the reading ability of various children who had been tape recorded earlier. What the teachers did not know was that the children were from the same five areas that they themselves came from. What was learned was that children were given low scores if they read in a dialect that was regionally or socially different from their teachers', that pauses counted as reading errors, that false starts and stuttering were stigmatizing, and that reading with expression counted in the evaluation of reading ability. If reading ability can be judged on the basis of the language clues related to regional or social dialect, it is possible that intelligence will also be judged in this manner.

In a study (Shuy et al, 1969) of Detroiters' subjective evaluations of people who used five types of speech (Detroit speech, White Southern speech, British speech, Negro speech, and Standard speech), 620 respondents were presented with 12 semantic differential scales. One such scale, using the polar adjectives *smart* at one end and *dumb* at the other, shows that people use a subjective assessment of intelligence as a conscious assessment of language. The seven-point scale was weighted as follows:

Smart ___1___ : ___2___ : ___3___ : ___4___ : ___5___ : ___6___ : ___7___ : *Dumb*

Thus, a person who thought users of Detroit speech were *very* smart would rank them in the slot numbered *one*. Likewise, if he thought such speakers *very* dumb he would mark the number *seven* slot. If he felt that they belonged somewhere in the middle he would rank them accordingly. From the four socio-economic classes of respondents, distinct patterns emerged. British speech was considered "smartest" by upper-middle-class respondents. British speech was considered increasingly "dumber" as the class of the respondent became lower. This pattern was exactly reversed in the responses to Negro speech and to White Southern speech. In terms of age, 10-to-12-year-olds and adults tended to respond similarly to people who use these five types of speech, whereas 16-to-18-year-olds were generally more negative about them all, considering them generally dumber. No significant differences were noted between the responses of males and females to these categories. Not surprisingly, Blacks ranked White Southern speech dumber than White respondents did and, likewise, Blacks regarded Negro speech smarter than Whites did. Blacks also had a lower respect for Standard English than did Whites.

If anecdotal evidence is permissible in our effort to identify areas of danger with respect to judgments about intelligence, the recollections of males who grew up in working-class communities may be called up as testimony. Men on whom I have tested this idea generally agree that one requirement for growing up male in a working-class community is to clearly maintain a sense of

masculinity. One way to be masculine is to be able to out-fight all other males. Another way is to excel in athletics. These are the major positive markers. Negative markers include avoiding any appearance of being a sissy. How sissy is determined is very complex and, as far as I know, unresearched, but in my day it included a favorable attitude toward school, and a general value for being marked in terms of school (or teacher) norms of intelligence. Boys who said they liked school were the lames in our society. Boys who did well in school were all right unless they displayed their success too obviously. That is, a male could do well on a report card or on a written test, but it was not well-thought-of to recite with high favor, or to display other evidences of surface intelligence. All this was negated, of course, if it should happen that the toughest male in the class was also the smartest. The prime condition of invulnerability, of course, was the tough who was also the top athlete. He was allowed to be as smart as he wanted.

I had forgotten all about this complexity from childhood until my teenaged son began showing evidence of reacting to similar expected norms in his school. Since our roles were different, the pattern became clear to me. Whereas I was about four on a one-to-ten scale of toughness, he is a 10. Likewise, he is a better athlete than I was and, as far as I can tell, we are about equal in intelligence (a fact which he will gladly dispute, but, since he is bigger, I will not contest). As a tough athlete, he is allowed to appear as intelligent as he desires, an opportunity of which he has, unfortunately, seldom availed himself. Of particular interest, however, is the reversal of our positions in another way. Whereas I was reared as an "I seen him when he done it" speaker of English, he grew up speaking almost exclusively standard English in a relatively cultivated community. Consequently, when faced with the role of playing defensive tackle on his football team, he had to develop new strategies for developing or reasserting his toughness. One surprising strategy was his cultivation of certain aspects of non-standard English, particularly its phonology. Now, as a teen-ager, he is finally mastering the art of unacceptable English in a way which was native to me. *These guys* is articulated as *dem guys* and *this* is realized as *dis*. One might speculate as to whether or not such a strategy is necessary for quarterbacks or flankers. I doubt it. But defensive tackles in an apprenticeship to become apes-in-residence must find the technique useful.

To this point it has been asserted that subjective judgments of all sorts, including judgments about intelligence, are made by teachers, employers, researchers, and the general public, regardless of the languages or settings in which they occur. Although clear evidence does not yet exist to show that minority speech codes are capable of expressing *everything* that the standard codes express, rather compelling evidence rejects every claim made by those who attempt to show linguistic correlates of cognitive deficit. The arguments put forth by educational psychologists such as Engelmann, for example, clearly misrepresent linguistic data and reveal naive field methodologies. One obvious area for improvement in educational psychology is in the development of better data-gathering instruments and analytical procedures. It is clear, however, that linguistics are in whole-hearted agreement that no one language or dialect, standard or non-standard, is known to be significantly more complex than another in its basic grammatical characteristics. Linguists have not found any speech community with a native language that can be said to be logically or conceptually primitive. Likewise, the so-called "non-standard" dialects of English spoken by lower-class families in the inner cities of this country are fully-formed, logical languages with only superficial differences from standard English. Despite all this, however, people tend to react to these superficial differences as though they were some clear form of cognitive inferiority.

Two questions remain for consideration here. Exactly what linguistic aspects are people responding to when they make judgments about a person's

intelligence based only on a sample of his speech? What aspects of their culture or society help lead them to such conclusions?

It is not at all clear what people respond to when they listen to language. The field of studying subjective reactions to language is still very new, and we are only beginning to understand some of its dimensions. In terms of grammar and phonology, it is becoming clear that listeners attend to stigmatizing features more than to favored ones. That is, stigmatizing features such as multiple negatives, consonant cluster reductions and *d* for *th* substitutions in words like *dis* for *this*, tend to count negatively against a person more than favored features such as clever wording or complex syntax which count positively. Perhaps this stems from a negative view of the world, a willingness to believe the worst about people, or from the same principle by which we recognize tact--by its *absence* rather than by its presence. Since research in Detroit speech has shown that residents of Detroit are far more accurate in their judgments about the ethnicity and socio-economic status of tape-recorded Detroiters of lower status than they are about Detroiters of higher status, it appears that stigmatizing features outweigh favored ones (Wolfram, 1970). It is difficult to prove that stigmatizing features that deal with grammar (such as multiple negatives) are more stigmatizing than features that deal with phonolgy (such as *dis* and *dat* for *this* and *that*), but all evidence would seem to lead in that direction. One clue comes from the fact that of the two types of features--grammatical and phonological--stigmatizing grammatical forms are lost faster as a person becomes more and more standard-English-speaking. If any linguistic features of non-standard are left in the speech of a person who has been rapidly increasing his standardness, they are apt to be pronunciations more than grammatical forms.

One interesting correlate of this condition involves hypercorrection, a term linguists use to refer to an incorrect overgeneralization from already learned forms. Several years ago, I noticed such a pattern in the development of my younger son's use of *-en* participles. Suddenly, he seemed to be using the inflectional *-en* in all participle slots such as *have thoughten, have senden,* and *have playen*. My first reaction was to drill Joel on the proper form, but I soon realized that he was actually evidencing awareness of a newly-acquired pattern. What he had not yet learned was how to sort the participles out into *-en* and non-*en* forms. That would take time, but it would come. Hypercorrection is perhaps more readily recognized by teachers in the form of the malapropism, a vocabulary item which comes close to the sound of the word intended, but which clearly misses, yielding a humorous combination such as "prosecuting eternity." Grammatical hypercorrection yields pseudo-elegances such as "between you and I."

Within each linguistic category (pronunciation, grammar, and vocabulary), individual features can be placed and rank-ordered, although the exact nature of this ordering is not totally known at this time. Such orderings should be done concurrently, however, for features which appear at the bottom of the list of stigmatizing features in grammar may count less than certain highly-stigmatizing pronounciations even though grammar, in general, stigmatizes more strongly than pronunciation.

But even data from hypercorrection can be misleading in any effort to make judgments about intelligence based on language alone. Research in child language acquisition shows that children tend to regularize the past tense of irregular (strong) verbs--*comed, breaked, goed, doed,* and so on. Regularization of this type continues into elementary school for some children. From a traditional psychological perspective, one might expect to find that children begin by using some regular (weak) forms correctly--like *walked* and *helped*--and that they then overextend this rule to the strong verbs. In actuality, however, the situation is different. In all of the cases which have been studied, the earliest past tenses are the correct forms of irregular verbs--*came, broke, went,* etc. Apparently these irregular verbs in the past tense--which are the more frequently

used past tense forms in adult speech--are learned as separate vocabulary items at a very early age. Subsequently, as soon as the child learns only one or two regular past tense forms, he replaces the correct irregular past tense forms with the incorrect overgeneralizations from the regular forms. Thus, children actually say *it comed off, it breaked,* and *he doed it*. The crucial point here is that the iregular verbs, though they are frequent, are each unique--they do not follow a pattern, and evidently it is patterns to which children are sensitive.

The schools have not generally taken advantage of this sort of overgeneralization either as an indication of an acquired stage in the development of acceptable school English, or as a positive indication that the speaker actually has to know something in order to produce such a form. The usual school attitude of correct or incorrect polarity toward error-making often discourages such insights. Mistake-making is seldom valued in the schools, and teachers soon learn to correct any errors that their children may make. This is incredibly short-sighted, since not all errors are alike, and many provide evidence of more creativity and cognitive ability than the presumed correct ones. My son evidenced such creativity once when asked where Australia got its name. He pondered a minute, reflecting that the country was settled by prisoners who were ostracized by the British, and explained that *Australia* was derived from *ostracized*. It was a creative answer which just happened to be wrong. The classic example of a virtuous error is the widely-told story of the physics student who, when asked how to measure the height of a building using a barometer, replied that he would go to the owner of the building and say, "If you'll tell me how tall your building is, I'll give you a barometer." Children experiment with language in much the same way. They try new combinations, they hypercorrect, they regularize irregular verbs, and they try many other creative and cognitively complex schemes. Perhaps the schools would do well to recognize different types of errors for what they frequently are--evidence of high intelligence. In any case, it should be clear that hypercorrection may well be stigmatizing in one sense, while in another sense, it provides evidence of creativity, intelligence, or the natural developmental process.

At this point, having asserted that people *do* make many judgments about speakers, including judgments about intelligence, based on a misconception of what language is and how it works, and having outlined some of the framework of the linguistic clues that are used to make such judgments, let us turn our attention to some of the reasons behind these processes of making judgments. It is my contention that there are at least three reasons involved in determining how these processes operate. For one thing, language is so close to us that we tend not to see it. Thus, we are easily led by stereotyping, which can be a useful way of organizing the universe, but which can also provide a very imbalanced and inaccurate view of reality. In addition to proximity and stereotyping, people do not understand language very well as an affective process.

It is clear that the early stages of learning in many areas of life require bringing to consciousness certain acts or knowledge which, having been acquired, quickly move to an unconscious state. In the early stages of learning to read, for example, it is said that a child calls attention to the letter-sound relationship in an obviously distorted way. That is, having once learned to decode sounds from letters, he henceforth pays little attention to this procedure. Otherwise, of course, he would be decoding units far too detailed and small for efficient reading. He apparently learns to decode language units of an increasingly larger and more complex state, and his reading involves language processing of morphemes, words, clauses, and sentences as he becomes a better and better reader. In fact, if the efficient reader should bring to consciousness those early stages of the learning process, they could conceivably hamper the operation at hand. Such a phenomenon may be illustrated by carefully considering what one should do with one's feet while climbing a stair, or by intellectualizing the intricate operations involved in driving an automobile while

negotiating in heavy traffic on the highway. Both could prove disasterous. In all skills, the goal is to become unaware of the parts of the process. In early life, man learns to talk, thus enabling himself to become conscious and, in a sense, splitting himself in opposition. Then he spends the rest of his life putting himself together again.

Time and again people can identify the races, socio-economic status, and general effectiveness of tape recorded speakers with amazing accuracy. But when asked to explain how they do it, they are usually at a total loss for words. They most frequently resort to stereotypes or hearsay, claiming that Blacks have deeper voices than Whites, or that Whites have more nazalization. Speech teachers tend to respond by muttering something about voice quality, but they almost never can explain what this means. Likewise, in the teaching of beginning reading, one technique--the "language experience approach"--requires a teacher to write down a story which a child tells. The theory is that the child will be able to read what he himself has uttered better than some primer story. However good the theory may be, it is practically never tested, for the teacher invariably writes down what she expects to hear, rather than what the child really said, often correcting the grammar and interpreting his hesitations.

There is no metalanguage for linguists to use to talk about language to non-linguists. We have been criticized for throwing our terminology around indiscriminately, but there is probably no other way for us to talk. All this makes language study an even richer field of investigation, of course, for it is possible that it is the best area for learning about man, without his catching on to what we are doing. Seeing language provides man with the same difficulty as a fish would have in seeing the water it swims in.

The thing we call "consciousness" is a kind of indoctrination which begins early in life, when the child learns a symbolic system which takes him out of the here-and-now. Learning the names of things is one step in this indoctrination, for as humans learn the names of objects, actions, and relationships, they come to live in a representational universe. The naming of things, or the representational use of names, implies a kind of stasis which seems to deny the movement which so characterizes life. In fact, language is really part of the schema of resistance--that which attempts to stop behavior at least long enough for us to conceptualize it and look at it.

This freezing of movement through symbolic language enables one to live in the mind (remembering the fact that symbols only approximate life at a moment in time), rather than to merely interact with the moving environment. In fact, the basic directions in the fields of philosophy, mathematics, and science (to say nothing of linguistics) are toward the representation of reality in a static fashion, frozen in time and space.

Thus, the representation of life in static terms, although necessary for human interaction about and with life, actually serves to mask the reality of its actual variety and change. In one sense of the term, at least, a word is only a rather poor stereotype of the meaning which underlies it. But we humans need even this stereotype, and so we focus on it and assume that it is real rather than a cartoon. It is of special interest to linguists that dichotomizing about language, whether it is an either-or statement about grammar or a less-comfortable generalization about phonology or semantics, often conveniently overlooks the fact that language operates on a kind of continuum which contains high degrees of variation. Grammars, by definition (at least in the past), have not dealt with variation in order to highlight large-scale patterns. The "grammar of a language" consists of these large-scale patterns. Of course, people do not conform exactly to such grammars in their actual linguistic performance. It is only recently that linguists, especially those who call themselves sociolinguists, have turned their attention to describing this variation--particularly the relations of linguistic variation to extralinguistic matters such as the style and intention of the speaker, and variables such as the speaker's status, sex, race, and age.

It is precisely at this stage of stereotyping where misconceptions about language affect judgments about intelligence. A major point is that we do not get to this stage only through ignorance or maliciousness but, quite the contrary, we are led to it very naturally by the paradox of our human condition. And it is indisputably true that people have many kinds of misconceptions about language, including those that affect judgments about the intelligence of the user of that language. The nature of our being requires us to correlate known facts with the unknown, however wrongly this process may lead us. The obvious corrective strategy is to study and learn what, if anything, language *can* tell us about intelligence, and to get a firm grip on what it cannot tell us. Very simply, I argue here for getting rid of misconceptions about language. The quickest and most efficient way to do this is to learn a great deal more about language than most of us now know. Several years ago, I argued for placing the study of language--what it is, how it is acquired, how it varies from group to group, how it changes, and how it relates to all other apsects of life--at the center of the training of teachers, rather than at the periphery. (The study of language should go hand in hand with the study of culture.) I have not been persuaded away from my belief that the study of language and culture should be at the center of teacher training by the passage of time, the advent of new programs, or the discovery of different educational management strategies.

If language stereotyping grows out of the natural process of language development, supplemented of course by the human potential for social vindictiveness, it would be well for teacher-training programs to give serious attention to how this stereotyping develops. In order to do this, we will need to raise the teachers' consciousness of the preceding stage--to make them aware of the language which is so taken for granted that they cannot see it. Thus, one may easily argue for educating teachers in how language works, how it is learned by children, how variation works in English, and how to use language to study and evaluate the language of others. All of this will lead to a better understanding of language affect than is now exhibited in the schools, and all of it will contribute to a more healthy assessment of how it is that misconceptions about language affect judgments about intelligence.

The Cultural Context of Learning to Read

R.P. McDermott,
Rockefeller University

In early research efforts, doctors looked for deficits in the sensory and neurological equipment of children unable to read. Recently, research has been conducted to locate impairments in the various perceptual, cognitive, and linguistic skills that children must develop if they are to become mature readers. This paper is different in that it will focus on the larger cultural context in which these component skills develop.

Specific ways to sensibly interpret and produce behavior constitute a culture. The procedures for producing reading behavior are only a small part of a person's culture. Reading behavior is only one of many different kinds of communicative behavior that children must learn from adults in order to become competent members of our culture. If reading is to be properly understood, reading behavior must be placed in context with all the other behaviors that children learn to produce.

The theme of this paper is that, for many minority or culturally different American children, reading failure is best explained by the cultural makeup of the classroom, and not by biological, psychological, or linguistic deprivation (Shuy, 1969). It is likely that many minority children do not develop the component skills essential to reading; their performance on standardized reading tests, at least, indicates that they do not. But the reasons for this illiteracy cannot be understood solely in terms of the component skills. Rather, reading skills, or the lack of them, must be understood in terms of the many other social skills that children develop. The *hypothesis* to be explored in this paper is that for the minority child in most American public schools, success in reading and success in social interaction with one's peers in the classroom are mutually exclusive. This is so because of conflict between the child's culture and the dictates of the teacher's culture. Even the best of teachers and the most intelligent of children can fall prey to such conflict, and together they produce failing reading records.

This "cultural approach" to reading failure complicates the research task by adding many variables. Nevertheless, a cultural approach may simplify our task in the long run. Consider the importance of cultural factors in the distribution of functional illiteracy in the American population: approximately half the members of most culturally and racially different minorities are unable to read (Thompson, 1966). This compares to a 10 percent rate of reading failure for White Americans and a one percent rate for the more culturally homogeneous populations of Scotland (Clark, 1970) and Japan (Makita, 1968).

Some have tried to account for these relations between social factors and reading skills by pointing to various deficits in the minority populations. Genetic inferiority, high rates of brain damage, and linguistic and cultural deprivation due to low social class have all been cited. But, all these theories

---

Editor's Note: *Some themes explored in this paper are explored further in: McDermott, R.P. "The Ethnocracy of Speaking and Reading." In* Linguistic Theory: What Can It Say about Reading?, *edited by R. Shuy. Newark, DE: International Reading Association, 1977.*

Ethnographic data relevant to these themes are analyzed in McDermott, R.P. "Kids Make Sense: An Ethnographic Account of the Interactional Management of Success and Failure in One First-Grade Classroom." PhD Dissertation [unpublished], Stanford University, 1976.

have been unable to account for the creativity of many failing children in social settings outside school. In addition, there are numerous other arguments against each deficit theory (Alland, 1971; Labov, 1969). This paper offers an alternative account. The attempt will be to show how reading failure may develop in even the most intelligent minority child if the child is educated in schools organized and staffed my members of the majority culture, regardless of the depth of the majority's concern for the child's development.

CULTURE AND READING FAILURE

The argument for culturally induced reading failure can take three directions. The first argument is obvious. In some cultures, reading has no functional importance. Knowing how to read will in no way enable a highland New Guinean child from the Fore tribe to better tend to his pigs and gardens. It is not difficult to understand why missionary schools are not too successful at teaching reading to the Fore (Rubinstein and Gajusek, 1970). Such an argument in no way applies to American minorities, for, in our complex society, literacy is a minimal requirement for access (Landau, 1971) to the information necessary for social mobility. Despite this, many minority group members remain illiterate.

A second argument claims that if the cognitive style of a particular culture is at odds with the skills necessary to reading, then the children of that culture will not learn to read. For example, it has been argued that some cultures are less lineal than others (Lee, 1950); some are less abstract than others (Cohen, 1969); some are less oriented to problem-solving than others (Gladwin, 1970). Members of such cultures supposedly have a harder time reasoning in terms of sequence, in terms of cause and effect, and in terms of past to present to future. Accordingly, it may be difficult for such people to produce reading behavior which is most often described in terms of "psycholinguistic guessing" and hypothesis testing (Hochberg, 1970; Wanat, 1968, 1972A, 1976A, 1976B).

But such arguments are inadequate at more general levels. The overall cognitive style of a culture has proven elusive and no adequate methods have been developed for delineating the different forms of logic employed in different cultures (Cole and Gay, 1972). Certainly all people play psycholinguistic guessing games and test hypotheses in their production of behavior. Members of all cultures are proficient at serially ordering their behavior and accordingly see, speak, walk, and interact. If members of some cultures appear to organize some aspects of their thought less lineally than others, there is no reason why reading should be any more affected than any other motor behavior. Culturally induced reading failures must develop from someplace other than the formal logic embedded in the culture's categories for action, for there are no data that indicate that any of the world's thousands of cultures logically disables its members from mastering reading skills once sufficient motivation and adequate presentation of the task are present.

The third argument for culturally induced reading failures constitutes the rest of this paper. The argument is complex. If a child's culture is at odds with the culture of the classroom in which reading instruction is offered, the child may not learn to read. If a classroom is divided into two incongruous ways of doing things (for example, White ways vs. Black ways or teacher's ways vs. peer group ways), and if reading is aligned with one particular way of doing things (the White teacher's way for example), then the children in the opposing faction may reject reading. As a consequence, the children *will learn how not to read.*

Two kinds of evidence support this attempt to account for the high rate of reading failure among minority students. First, there is an increasingly large amount of evidence that when a dominant group within a society attempts to educate minority children, with almost no exceptions, the minority children fail in

school. We need to look no further than to 300 years of White failure in the education of Native Americans and at least a century of failure in the education of Blacks and Chicanos to give this argument some force. The Japanese have fared no better in their attempts to educate the Koreans (Mitchell, 1967), or the Burakumin (DeVos and Wagatsuma, 1966; Shimahara, 1971). One might also cite the education of minorities throughout the world, the Lapps in Norway, Native Americans in South America or Irish Catholics in Belfast.

This record of failure contrasts with the successes of minority groups in charge of their own schools. Such groups are successful in transmitting the values of their own culture to their pupils; they also produce successful students as measured by the standards of the majority school system. The Pennsylvania Amish offer an outstanding example. In public schools, they suffered school failure and the degradation of their culture. With the establishment of their own schools, the children were nurtured in their traditions (Payne, 1972) and achieved some of the top reading scores in the state (Hostetler and Huntington, 1971). Tibetan and Mongolian populations in China are educated in their native tongue and according to local customs. Chinese characters are gradually introduced as the graphemic dimension of the native tongue. The switch to eventually reading and speaking Chinese is apparently a smooth one (Allitto, 1969). American Catholics must have intuitively understood this Chinese strategy. By setting up their own schools and granting their own diplomas, they have neatly fit into American institutions.

Apparently, when the classroom can be spoken of in terms of a community of shared meaning and shared codes for expressing that meaning, children acquire whatever codes the classroom has to offer, be they linguistic, religious, or educational. This postulate is well illustrated in the case of the German classroom in which the teacher often speaks High German and the children Low German, a difference apparently far greater than the difference between Black and White American speech. The German dialect difference, however, apparently makes little difference as the teacher conducts the class in Low German (Fishman and Leuders-Salmon, 1972). Speaking and reading High German becomes a game at which both the teacher and the children play. The question is whether a child's own communicative style of the classroom gives the child the incentive to communicate with his classmates and teachers in whatever codes are available.

A second line of evidence for reading disabilities being caused by culture conflict is more specific. First, it will be shown that selective inattention to certain kinds of stimuli is developed in everyday life in response to a child's experience with the stimuli in question. Second, an account of Black children acquiring selective inattention to reading materials is offered. Third, this selective inattention is explained in terms of culture conflict.

CLASSROOM POLITICS AND SELECTIVE ATTENTION

Attention is a mechanism that helps us to carry out the division of the world into significant and insignificant. As Wallace (1965) has suggested, man has created a rich and elaborate world for himself "by a process of selective attention to his total environment." Apparently, man suffers tremendous limitations in his ability to process all the information available to him at any one instant. Selection is a ubiquitous and unending process throughout our central nervous systems. Significance is stored not only in the social world, in the symbols out there, but in our equipment for decoding and interpreting the world. What is organized in cultures and social organizations are individuals interpreting the world. What is organized are human brains attending to some aspects of the social world and not to others.

The same symbols that are processed by our bodies are the symbols that are processed in our cultures. Pribram (1971) has given one delightful example: "For many years there was an elevated railway line (the 'el') on Third Avenue in New York that made a fearful racket: when it was torn down, people who had

been living in apartments along the line awakened periodically out of a sound sleep to call the police about some strange occurrence they could not properly define. The calls were made at the times the trains had formerly rumbled past. The strange occurrences were, of course, the deafening silence that had replaced the expected noise." A model of the world for a particular time and place existed in the central nervous systems of many people up and down the Third Avenue El. When the environment did not supply the expected information, the people began paying attention. In this case, the technology which had worked its way into our bodies told us when to listen, namely, at the time when the El had previously roared by; the culture which had worked its way into our bodies told us what to hear when we listened, namely, a thief or another strange occurrence requiring police assistance. Culture and technology are merely two aspects of the environment with which we communicate. They send us messages about the adequacy of the internalized models in terms of which we perceive, think, and act. We all live in a world of information which we decode according to the dictates of context. Much of this context is encoded in our memory and evidenced in our attention patterns.

Half a century ago, attention was a major focus of classroom research. This research attempted to document whether or not children were paying attention to their teachers by noting where they were looking during lessons. This literature documents contrasting attention patterns between middle American "host" (majority-culture) schools then and minority classrooms now. It was shown that more than 90 percent of the majority-culture children of the 1920s had their eyes fixed on their teachers or their work at any given time (Jackson, 1968). This contrasts considerably with estimates obtained from the contemporary classrooms that share in the early century style of the teacher directing all attention in the classroom. Deutsch (1963) has found that teachers in Harlem elementary schools spent more than half their day calling children to attention. Attention patterns indeed appear to define the "scene of the battle" in minority group education (Roberts, 1970).

In the classroom where teachers and students produce leadership patterns for each other, attention is an issue. To attend to a teacher is to give the teacher a leadership role in the classroom; to attend to the peer-group is to subvert the teacher's role. In the older studies, the primary fact is that all the children paid respect to the teacher's leadership role by attending, physically at least, to the teacher's activities. In schools populated by minority children, this leadership role is much more subject to negotiation: some teachers succeed at this and some do not; some children give their attention and others do not. It is in the context of this battle for attention that we must consider the nature of minority reading failures. In many intercultural classrooms, the "politics of everyday life" has been escalated into "war games." Children in some groups fight the teacher by misbehaving whereas others choose to build a wall of silence and nonresponsiveness in order to exclude the teacher (Dumont, 1972). The particular war strategy is unimportant. What is important is that there are teacher games and peer group games, and every student must make his choice. One takes sides by attending or not attending. Those who attend learn to read; those who do not attend do not learn to read.

LEARNING NOT TO READ

Attention patterns increasingly shift from teacher games to peer group games as a minority child moves through elementary school. In addition to the often reported facts that minority children learn less and misbehave more as they get older, data on shifts in the children's perceptual capacities, their language, and their attitudes are now available.

Perception

Learning how to read involves a great deal of perceptual learning. Many

children reverse their letters when first learning how to write, for there really is not much difference between *p, d, g* and *b; o* and *e; u, n, m, v* and *w*. Most children master these subtle differences, but disabled children continue to have difficulty distinguishing these forms after the first grade. Careful attention must be paid to differences. We develop these skills and store them deep in our nervous system. The more we read, the less work it takes to distinguish the different forms. The eye apparently learns just what to look for and orients only when a drastically misshapen form appears; the difficulties of proofreading illustrate just how well a good reader is programmed to read for meaning and to notice typographic irregularities only when they are given special attention.

Some Black children do not permanently develop the essential skills for letter differentiation. In a test designed to analyze a child's competence at handling the perceptual transformations essential to letter discrimination, Gibson (1965) has shown that most children have trouble with certain transformations. However, Black children I tested at 12 years of age showed a mixed range in these skills; those who could read performed well, and those who could not read performed very poorly, scoring below younger Black children on the same test. These Black children had apparently learned how not to see. Reading apparently became a call for inattention.

Language Structure and Function

The work of Labov and his associates in Harlem has revealed that our language is socially organized. The way in which our vocal chords allow for a passage of air to reverberate into the ears of other social actors depends greatly on just who the interactants are and how they are related. Depending upon who is doing the talking and who is doing the listening, not only will different points of information be passed, but the way of saying it may be remarkably different. Minority children often learn to use one speech code or "register" for dealings with minority people and another for host people; the differences in the code may be the unique usage of particular words as in Japan (Donoghue, 1971), subtle sound shifts as between social classes and ethnic groups in New York City's Lower East Side (Labov, 1964A), gross dialect shifts as between Whites and Blacks in American, or major language shifts as between French and English speakers in Canada, or the Lapps and Norwegians in Norway (Eidheim, 1969).

American Blacks acquire a dialect of English that is most often mutually intelligible with White English. When there is an intelligibility breakdown, the result can be disastrous. Gumperz (1970) gives the following example from a post-bellum Southern teacher's diary: "I asked a group of boys one day the color of the sky. Nobody could tell me. Presently the father of one of them came by, and I told him of their ignorance, repeating my question with the same result as before. He grinned: 'Tom, how sky stan'?' Blue, promptly shouted Tom." This gross level of language interference was, of course, attributed by the teacher to the child's stupidity, and the teacher probably unconsciously related to the child the subordinate status that accompanies being "stupid" in a classroom. For the most part, however, Black and White verbal codes are mutually intelligible in content. Switching does not cause a problem at the structural level; rather the codes differentiate the games being played and the meaning of the behavior displayed by various players in the game. Messages of relationship are differently stored in the codes, and, by subtly manipulating aspects of the codes, a speaker can inform persons just how they stand in the speaker's hierarchy of persons. Using one code and not another is itself a powerful relational message. It is not difficult to learn that "What color is the sky?" and "How sky stan'?" are equivalent, but, when your teacher deems you ignorant for using the other, then the job of switching codes is difficult indeed.

Black peer group games are well defined by the growth of a highly elaborate linguistic code restricted to peer group members. Labov (1964B) has

isolated these games in his delineation of the stages of acquisition on non-standard English: 1. Up to age 5: basic grammatical rules and vocabulary are taken from parents; 2. Age 5 to 12, the reading years: peer group vernacular is established; 3. Adolescence: "The social significance of the dialect characteristics of his friends become gradually apparent;" and 4. High school age: "The child begins to learn how to modify his speech in the direction of the prestige standard in formal situations or even to some extent in casual speech." The second and third stages are most important for considering the implications of the school and peer group speech codes for learning. The implications are not so obvious; what difference does it make if children use one register for interacting with teachers (and reading materials), and another for interacting with each other? The important thing is that these two codes become mutually exclusive during the school years. As the child participates in the culture defined by his peers, the more deviant his linguistic registers. It is these linguistic features which help to mark off the peer group from the culture of the schools.

Participation in peer group formal organizations and the employment of their linguistic registers are of great importance, for they correlate strongly with reading scores (Labov and Robins, 1969).

Attitudes

Labov's findings are not limited to classrooms in Harlem which harbor members of formally organized gangs. The same trends were found in two classrooms I studied in suburban New York City. The children were not aware of the trends and their significance was readily apparent in the children's attitudes toward each other. This should not surprise us at all. In my high school, most of us could define others and make very accurate estimates of their grade point averages on the basis of their clothes, speech patterns, and some postural cues. Our expertise was perhaps not as loaded as that of Black children, for their expertise not only defines others but also determines who is to be popular or not. A series of sociometric tests administered in an all Black, bottom-track, sixth-grade were consistent in placing non-readers at the center of all peer group activities. Similar tests in an all Black, non-tracked, fifth-grade also showed non-readers at the center of most activities. Reading skills do not recommend an actor for leadership. Indeed, the acquisition of such skills can exclude an actor from the peer group culture.

CULTURE CONFLICT AND READING FAILURE

The hypothesis of this paper is that reading failures may be products of the way that people in the classroom interact to produce statuses, identities, and ways of attending to stimuli in the classroom setting. Minority children do not have to be intellectually inferior, brain-damaged, or deprived to fail miserably in intercultural classrooms. Teachers do not have to be racists for the intercultural politics in the classroom to produce a setting rigidly divided between teacher and peer group games. Any formal difference in the communicative styles of the teacher and the children can introduce havoc to their relations and the messages of relationship they consequently send to each other.

Communicative-code differences in a classroom setting can have tremendous effects. Learning is most possible when both the students and the teacher share procedures for organizing time and space in the classroom (Collier, 1973). A teacher out of phase with his students will undoubtedly fail in the politics of everyday life. Rational interaction with the group will hardly be possible. As a result, the teacher will fall back on his formal authority as a teacher, his so-called "role," to instruct the children in their classroom behavior. The children often reject this authority role and develop an idiosyncratic code, such as the non-standard peer group code Labov has described. The children's actions make so much sense. When rational interaction with a teacher is not possible,

i.e. when his position of authority makes no sense in terms of his relations with the children, they produce an alternative system and disown the teacher's authority. Reading skills get caught in this battle over which cues are to be attended to--peer group cues or teacher cues.

Communicative Code Differences and the Supression of School Learning

Spindler (1963) studied the self-fulfilling prophecies of teachers unconsciously dominating classroom social organization. He showed that middle class teachers attended to middle class children and labeled them as the most talented and ambitious children in their classes. School success followed along identical lines, but more subtle evaluations of talent divided the populations along different lines. In this case, lower class children gave up trying and acquired failing "institutional biographies" (Goffman, 1963) because of an inability to give evidence of their intelligence in terms of the limited code that the teachers used to evaluate children.

The same point has been r[illegible] for social class differences between Black teachers and Black children i[illegible]t. Louis. Rist (1970) offers the following account of the classroom after it had been divided into three "ability groups," the fast, slow, and non-learners at Tables 1, 2, and 3, respectively: "The organization of the kindergarten classroom according to the expectation of success or failure after the eighth day of school became the basis for the differential treatment of the children for the remainder of the school year. From the day that the class was assigned permanent seats, the activities in the classroom were perceivably different from previously. The fundamental division of the class into those expected to learn and those expected not to permeated the teacher's orientation to the class."

Assignment to each of the tables was based on the teacher's subjective evaluations which, after dissection by Rist, were shown to be rooted in the teacher's evaluation of the children's physical appearance and interactional and verbal behavior. Children with newer and cleaner clothes and more of them on cold days, slightly lighter skin, and processed hair were at Table 1. Children with reciprocal traits were positioned at lower tables. Class leaders or direction givers also clustered at Table 1. The children at the low tables spoke less in class, in heavy dialect when they did, and almost never to the teacher. What is most unfortunate is that by the third grade the children at the lower tables were still at the lower tables. Once the child is tracked, it is almost impossible to break loose. The lower the table, the less the child gets in instructional time. In addition, teacher expectations follow the child from year to year. Apparently, the acquisition of a school biography is completed within the first week of school, based on a teacher's ethnocentric evaluation of a child's mannerisms. The children in an intercultural classroom have few choices. They can take the peer group as a source of identity and fight the system; thus, the children at Tables 2 and 3 potentially transformed by late elementary school into the gangs Labov has described. Another and perhaps worse choice is represented by the children at the lower tables who accept the teacher's definitions and passively fail through school and fall into pariah status. They also fail in establishing identity. For the children to dispute the messages of relationship offered by a teacher to the lower ability groups causes havoc in the classroom but solid ego development in the children's own community. For the children to passively accept subordinate status creates classroom calm, but a weak ego. Either way, learning is blocked; in the first case by active selective inattention and misbehavior, in the second case with motivational lag and selective inattention. Neither group learns to read.

Black-White Communicative Code Differences

Blacks and Whites slice up the world in slightly different ways. When the Black child enters a host school, he is asked to drastically alter his codes for

spacing his body vis-a-vis others (Aiello and Jones, 1971; Hall, 1971; Scheflen, 1971, 1972) and his timing in conversation (Lewis, 1970; Leonard, 1972; Gumperz and Herasimchuk, 1973). Punctuation breakdowns may result in the self-fulfilling prophecies described above. If the child moves and speaks in the wrong way at the wrong time according to the teacher's code, then he may be branded hyperactive, out of control, or stupid. The teacher will appear equally disoriented according to the child's code and may well be branded cold and unfair. Slight differences in the use of time and space do not have to result in such disaster, but they often do.

Byers and Byers (1972) have described how Black-White code differences can lead to pupil-teacher communication breakdown. The teacher in question is considered unbiased and talented. Byers and Byers filmed her interaction with four-year-old-girls, two Black and two White. In the sequence analyzed, she looks at the Black and White children almost an identical number of times. However, the Black children look at the teacher more than three times as often as the White children. One might postulate that the children are anxious about their performance, or perhaps they are hyperactive, etc. However, a far more interesting hypothesis is put forward. Byers and Byers found that the White children established eye contact with the teacher almost twice as often as the Black children who are straining three times as hard to catch the teacher's eyes. This is very crucial, for it is during eye contact that the teacher can send the children messages of reassurance and affection, messages such as, "I love you," "You are doing well," "How smart!," or at least, "You are making sense to me." Interaction is an action between people, and if not perfectly timed the interaction will fail. In the case of this teacher and these children, the White-White interaction is successful, and the White-Black interaction fails. For all their interactional work, the two Black girls receive little relational support, and it is not difficult to predict that they will someday direct their interactional work toward each other. In time, achievement will be located in the peer group and not in the teacher. The acquisition of two cultures, or biculturation (Valentine, 1971), can have its drawbacks for ego development. The bicultural child must acquire two sometimes mutually exclusive ways of knowing how to act appropriately, one way for when Whites are present and another for when the interaction matrix is all Black. Where code shifting is most difficult is apparently in the bureaucratic setting in which the White code, in addition to being the only acceptable medium of information exchange, is also the medium for the expression of host group power and host group access to the essential and even luxurious utilities of contemporary America. In the classroom, the teacher has the power; the teacher has the tools to supply the institutional biography that the child needs to insure upward mobility. Teachers are quick to point this out to children and daily tell them that there is no success without school. If the children attempt biculturation, they adapt to the teacher's code, accept the teacher's messages of power and dominance, and work hard at school. Many Black children do not go this route. In making their code make a difference, they are learning how to produce pariah status vis-a-vis the host group; they are learning to appear like "one-of those"; they are producing a pariah biography which will haunt them until the next generation again plays the politics of intercultural life in the classroom. (The implications of this stand for an account of social stratification in America are developed in McDermott [1974]).

SUMMARY

Three possible arguments for the influence of culture on high rates of illiteracy among the world's minority groups have been considered. The first argument claims that literacy will not be attained if it is functionally unimportant in the daily lives of the children. This may be interesting in an African or similarly underdeveloped setting, but the argument is irrelevant to reading failures in any industrialized nation with efficient mass communication

facilities. Exploration of a second position, namely that some cultures may better equip their members with the logical tools essential to the reading task, presents severe methodoligical difficulties and a lack of supporting data.

The third argument was discussed in more detail. Essentially, it is claimed that minority reading failures can often be understood as rational and motivated adaptations to communicative conflicts between minority children and teachers from the host culture. If a child's culture and the teacher's culture are different enough that they sometimes are unable to make sense of each other, the classroom will become a battle ground for the two opposing ways of knowing about the world.

Whether or not a child attends to reading materials depends upon the child's experiences with those materials, who gives them to him, when, and why, and how he feels about himself after using them. If the class is divided by a culture conflict between most of the children on the one hand, and the teacher and a few children, on the other, reading materials will be introduced to the child by the wrong person, at the wrong time, for the wrong reason, and the child will develop bad feelings and selective inattention to reading. Black American children in America learn *not* to read. Peer group membership, as measured by dialect skills and sociograms, is virtually mutually exclusive of literacy in the population studied by Labov. If a Black child is to be a leader in the everyday life of his contemporaries, then literacy is not a prerequisite; indeed it may be a handicap. Cultural conflict can inhibit school learning and lead to children learning how not to read. Communicative conflicts can easily develop between Whites and Blacks and the suppression of literacy by many Black children in White schools accordingly makes sense as an adaptation to everyday life in the intercultural classroom.

Criteria for Evaluating Readiness: "When Is a Child Intelligent Enough to Read?"

Stanley F. Wanat,
California State University, Fullerton

The question, "When is a child intelligent enough to read?" lies at the heart of much of the inquiry into reading readiness. Indeed, the goal of many reading readiness assessments is to answer the question: "Is this child intelligent enough to begin learning to read at this time?" This paper deals with procedures for assessing the linguistic-cognitive-contextual prerequisites for making easy progress in the acquisition of reading skills.

One approach to a discussion of procedures for testing reading readiness would be to begin with a definition of reading readiness which, in turn, must be based upon a detailed description of reading processes. Unfortunately, this approach has several pitfalls. Since reading is the most researched area of educational curricula, one might expect that reading researchers agree about what reading is. However, perusal of the reading literature will reveal that this is not the case. (See, e.g., Davis, 1971; Geyer, 1972; Kling, 1972; Wanat, 1971; Williams, 1973.) Furthermore, one needs to bear in mind that the skills present in *mature* reading may not be directly translatable into a reading *readiness* assessment: intermediate levels of skill acquisition may be present in learning-to-read processes. In this case, it might be necessary to relate testing procedures for assessing reading readiness to these intermediate skills, rather than to the skills involved in mature reading.

Gibson and Levin (1975) discuss reading readiness by considering the basic skills a child must have acquired for even the initial stages of reading instruction to be successful. They point out that a number of subordinate skills must have developed to some level of maturity before a child can benefit from actual reading instruction. Their exploration of this issue yields the following major components of reading readiness: (a) the ability to discriminate speech sounds, (b) development of graphic discrimination and writing, (c) intermodal matching and association, and (d) deeper cognitive abilities--perceptual, attentional, conceptual, and linguistic. Downing and Thackray (1971) define "readiness activities" as "a means of narrowing the gap between the state of the human individual and the conditions involved in learning to read." Issues involved in measuring that gap are the subject of this paper. The first section of this paper deals with the question: "Why test?" This section notes and discusses some of the purposes of testing. The use of single- versus multiple-factor criteria to assess reading readiness is discussed in subsequent sections, and concrete guidelines for understanding and utilizing reading readiness assessment procedures are provided in the concluding portions of the paper.

WHY MEASURE THAT GAP?--AND, MORE GENERALLY--WHY TEST?

Why test? The answers to this question indicate a wide variety of opinions about the purposes that tests can serve. J. Mackworth (1974B) discusses three main purposes for testing: "The first is *[#1]* to compare a group of children such as a particular grade with their peers in that or other schools. The majority of tests appear to conform to this purpose....The second purpose is *[#2]* to discover *whether* a particular child needs extra assistance, and if so, *[#3] where* his difficulties lie in the chain of mental activities from visual input to spoken output." N. Mackworth (1974B) notes that "existing reading achievement tests involving pencil-and-paper do seem to have a place in *[#4]* screening for the 15 percent of the school population who are reading disordered." Hoover (1974) states that "achievement tests are necessary in order *[#5]* to make the schools accountable

to parents." It follows from Hoover's statement that tests may enable and other members of a community to determine how well their children are learning in relation to children in other schools and communities. The goal of accountability (purpose #5) is served by the group-comparison function of tests mentioned as purpose #1. The appropriateness of holding the "educational delivery system" accountable needs to be considered with regard to teacher training programs as well as with regard to teachers and the schools (Wanat, 1972B).

Related to the accountability function of tests, testing may have the additional purpose of *[#6]* providing an "outside opinion" about a child's progress or potential, thus providing some protection from a situation in which circumstances lead a teacher to misjudge a child. Teacher expectations about a child's future level of academic success are often made very early in the teacher's contact with the child, and they may be based upon minimal information. McDermott (1977, research in progress) has observed that teachers differentially attend to and interact with groups of children which differ in their expected levels of success. Since teacher expectations can also affect the child's actual subsequent success (Brookover, LePere, Hamachek, Thomas, and Erickson, 1965; Henderson and Long, 1971; Palardy, 1969; Purkey, 1970), the availability of an outside appraisal of a child's achievement and potential might in many cases prove an asset.

It should be noted that Hoover sees the goal of accountability as being served by reading *achievement* tests; she is not in favor of reading *readiness* tests. One of her objections is that readiness tests tend to "sort" children into groups. However, since economic constraints usually rule out individual instruction, instruction is carried out on a group basis. The effectiveness of the match between the instruction and the needs of the children depends upon the homogeneity of the children in the instructional group. "Sorting" children for instruction is a central purpose of reading readiness testing.

A second purpose for reading readiness testing is to identify the children who may need extra assistance. In practice, this identification would result in a particular kind of instructional grouping. N. Mackworth (purpose #4) believes that available paper-and-pencil readiness tests can be used to screen "the 15 percent of the school population who are reading disordered." J. Mackworth (purpose #2) assigns a somewhat broader purpose for testing--determining whether a particular child needs extra help. Although J. Mackworth is opposed to reading *readiness* tests, readiness tests, in theory, could be used to perform this sorting task. In actuality, available reading readiness tests have many limitations which restrict their usefulness for this type of sorting.

A third purpose for reading readiness testing is to determine the locus of a particular child's difficulties. This purpose is least satisfactorily served by currently available tests. Some reading readiness tests may be able to predict, to a certain extent, a child's later reading achievement. However, reading readiness tests generally lack diagnostic power--they cannot isolate specific areas of weakness whose remediation would then facilitate the child's progress.

Thus, major purposes for reading readiness assessment include (a) assigning a student to a particular instructional group, (b) determining which students need remedial assistance, and (c) identifying a child's specific difficulties. All of these purposes are directly related to instructional decision-making. The view that one purpose of reading testing is to hold the schools accountable for their level of performance does not pertain to reading readiness testing, since such testing generally takes place when children first enter the school system. However, the growth of preschool educational programs may lead to the use of reading readiness tests to evaluate the success of particular preschool programs in preparing their students for reading instruction. The related role of readiness tests for providing an outside check on teacher judgments of a child's potential is another reasonable and important purpose for readiness testing. The use of reading readiness tests for diagnosis, while desirable, cannot be adequately accomplished with currently available readiness tests.

A SINGLE-FACTOR APPROACH

In a detailed historical analysis of the concept of reading readiness, and of how it has been assessed, Durkin (1970) states that "traditionally, readiness has been viewed as a product; for instance, in the beginning, readiness to read was assumed to be the product of maturation." "Maturation" in this case was the single-factor criterion used to gauge readiness. Both chronological age and mental age have been used as indices of maturity, and thus of reading readiness. The question of reading readiness was thought to be a simple yes-or-no, all-or-none question. If the child had reached a certain chronological (or mental) age, then he was ready to learn to read. If not, he was not yet ready. The instructional implications of this approach to reading readiness are obvious: if a child has reached the criterion age, then teach him to read; if he has not, then don't teach him to read.

The conception of reading readiness that forms the basis for this assessment strategy is an extremely simple one: reading readiness is the single product of a single process (maturation). Durkin (1974) critically evaluates the use of a single-factor criterion for determining reading readiness: "What about the frequently reported idea--even in current texts--that a mental age of 6.5 years is necessary for success with reading? This particular notion...is closely associated with a study of first grade reading described by Morphett and Washburne in a 1931 article, 'When should children begin to read?'. Central to their research was a particular kind of first grade reading instruction given in a particular kind of setting; and for success the combination appeared to require a mental age of about 6.5." Thus, the findings of Morphett and Washburne, obtained with "a particular kind of first grade reading instruction given in a particular kind of setting," were overgeneralized. In a study of the use of mental age as a measure of readiness, Gates and Bond (1936) found that "correlations of mental age with reading achievement at the end of the year were about .25. When one studies the range of mental ages from the lowest to highest in relation to reading achievement, there appears no suggestion of a crucial or critical point above which very few fail and below which a relatively large proportion fail."

Weintraub (1967) discusses the usefulness of reading readiness assessment procedures: "The most commonly used predictive measures of success in learning to read have been readiness and intelligence tests....Readiness tests tend to correlate somehwere between .4 and .6 with later measures of reading achievement, while intelligence tests, for the most part, show an even lower relationship at the lower reading levels. The readiness tests do an adequate job of identifying the extremes on the normal curve, those who will probably succeed and those who will probably fail. However, the large group of children in the middle may go in either direction when placed in a reading program." Gibson and Levin (1975) take the same position. MacGinitie like Durkin, rejects the use of any single-factor criterion to gauge reading readiness. MacGinitie (1969) succinctly defines reading readiness: "The concept is a very simple one: The child is in school to learn--what and how is he ready to learn? The notion of readiness is no more than that."

Within the framework of this definition, assessment of a learner's reading readiness must be made in terms of the skill(s) he is to learn, and the assessment must also take into account how the child is to acquire those skills. In accord with this definition, *reading readiness assessment is a process of gauging the match between learner characteristics and task demands.* For the most part, MacGinitie's concern for the question of "*How* is the child ready to learn?" has not been incorporated in readiness assessment procedures, for, as Rude (1973) notes in his review of five reading readiness tests, "Attention span, cognitive learning style, and experiential background are only three important factors which are not measured in the five batteries [reading readiness tests] examined."

The findings presented above argue that, in general, one cannot use such single-factor measures with confidence. Furthermore, the evaluations of Weintraub and of Gibson and Levin argue against placing much confidence in the use of currently available reading readiness tests.

MULTIPLE-FACTOR APPROACHES

The variety of viewpoints about the nature of the skills that make up reading is equalled by the diversity of opinion regarding which skills are prerequisites for learning to read. As a result, the abilities tapped by one reading test differ from those measured by another. The same is true for reading readiness tests. Rude (1973) notes that: "There is lack of consensus among test authors as to which [reading readiness] skills should be assessed as well as the techniques employed to assess them. The Gates-MacGinitie Reading Test--Readiness Skills, for example, includes seven separate subtests while the Murphy-Durrell Reading Readiness Analysis consists of only three subtests." This lack of consensus suggests that it might be preferable initially to consider a broad range of factors potentially involved in reading readiness, rather than trying to define reading readiness in terms of an extremely limited set of factors, or indeed, in terms of just one factor, which was a dominant approach in the past.

An analysis of eight reading readiness tests used in the U.S. is discussed by Downing and Thackray (1971). It was found that "of the eight, all use a test of visual discrimination, six use tests of vocabulary, three use motor tests, two use tests of the reproduction of patterns and shapes from memory, and two make use of tests of relationship. Other tests used include: ability to recall a story, ability to remember ideas in sequence, pronunciation, rhyming of words, auditory discrimination, and handedness and eyedness."

On the basis of an analysis of five reading readiness tests (Clymer-Barrett Prereading Battery; Gates-MacGinitie Reading Test--Readiness Skills; Harrison-Stroud Reading Readiness Profiles; Metropolitan Readiness Tests; and Murphy-Durrell Reading Readiness Analysis), Rude (1973) identifies 12 subskill categories comprising readiness: (a) vocabulary knowledge, (b) listening comprehension, (c) letter recognition, (d) numerical concepts and operations, (e) visual-motor coordination, (f) determination of whether or not two words rhyme, (g) phoneme correspondence (Rude expresses surprise that only three out of the five readiness tests he analyzed included a subtest dealing with this category), (h) rate of learning (only one of the five readiness tests considered by Rude contains this subtest, which provides an indication of the ease with which a child learns to read a sample set of words), (i) sound discrimination, (j) blending individual sounds together, (k) word reading, and (l) selecting from a visual array the visual form that matches a sample.

In an appraisal of selected research on reading readiness, Rude (1973) identifies four categories of prereading skills: (a) grapheme perception, (b) left-to-right visual scan, (c) grapheme-phoneme relationships, and (d) phoneme blending. An analysis of skills involved in reading readiness (J. Mackworth, 1974B) provides information on skills involved in the process of reading acquisition, as well as information on skills necessary for reading readiness. She states that "the necessary prereading skills involve adequate sensory discrimination, particularly in the visual and auditory areas. But in addition, the child needs a good spoken vocabulary that is consistent with the kinds of reading material that will be presented to him, and a good knowledge of grammar together with an understanding of word order, as well as good comprehension. All this depends greatly on his early environment....[Another] prerequisite is a good knowledge of his letters....In addition to the prereading skills of language and letter naming, the young reader has to learn a whole new series of skills, involving visual memory, parallel processing, and the recognition of left-to-right order of letters and words. Beyond these visual skills, he must learn the sounds made by the written symbols, and how a wide range of written groups of

symbols can have the same sound. Finally, he needs to master his visual behavior, moving his eyes from left to right along the type, and fixating on the important words within each sentence. He must master a reasonable speed of reading, because if he is too slow, he will have few words in his visual store, and he will read like a caterpillar, humping from word to word. His comprehension span will be too short for him to gain an understanding of the whole story, and his ability to predict what word comes next will be impaired." While other skills could be added to make Mackworth's description complete, she provides a useful listing of skills necessary in reading readiness as well as in learning to read.

While there may be factors that could be added to this description, it provides for useful identification of skills in reading readiness and skills that need to be acquired in the process of learning to read. Examination of the preceding description yields the following enumeration of factors in readiness: (1) comprehension of spoken language; (2) knowledge of grammar; (3) spoken vocabulary; (4) understanding of word order; (5) ability to predict the next word; (6) auditory discrimination; (7) knowledge of letter-sound correspondences; (8) knowledge of the symbol-sound relations in which (a) a given sequence of letters may be associated with a number of different sequences of sounds, or (b) a given sequence of sounds may be associated with a number of different letter sequences; (9) parallel processing skills; (10) recognition of left-to-right order of letters and words; (11) visual discrimination; (12) visual memory; (13) letter knowledge; (14) left-to-right visual scanning behavior; (15) selective visual attention to the important words within sentences; and (16) speed of reading.

In a "how-to" article for parents which provides suggestions for helping children develop readiness for reading, the following subskill categories are mentioned (Nelson, 1972): visual discrimination, visual memory, auditory discrimination, auditory memory, general knowledge, verbal expression, muscular control, and sense of responsibility. Consideration of the Downing and Thackray, the Rude lists of subskill categories in extant reading readiness tests, and the Rude, J. Mackworth, and Nelson lists derived from research on reading readiness, acquisition, and achievemnent provides evidence that an overall consensus on what abilities should be assessed in reading readiness is lacking.

Examination* of research on linguistic-cognitive-contextual prerequisites for reading identifies the following factors: (1) child's present level of reading performance; (2) ability to understand and follow the language of the directions to be encountered in the readiness assessment and in the instructional program; (3) physical efficiencies that might obstruct learning: (a) auditory acuity, (b) visual acuity, (c) other physical efficiencies; (4) level of comprehension of spoken language, including factors such as vocabulary development and concept development; (5) linguistic awareness: (a) understanding that print is related to language and to speech, (b) sensitivity to language sounds (e.g., rhyme), (c) ability to deal with words and sounds as objects; (6) learning skills: (a) proficiency in all language variables to be encountered, (b) attention span, (c) modality strength, (d) learning rate; (7) decoding skills, facility with: (a) letter discrimination, (b) letter order within words, (c) sound-spelling correspondences, (d) spelling patterns, (e) syllabic units, (f) structural analysis, (g) sight-word recognition, (h) applying his knowledge of syntax and semantics; (8) child's attitude: (a) toward different language varieties as media of instruction, (b) toward school, (c) toward learning to read; (9) teacher's understanding: (a) of the child's language, (b) of the child's culture, (c) of the child's

*Identification of these factors is based upon the deliberations of a subgroup at a working conference on the evaluation of reading readiness. The conference was organized by the present author, and the subgroup was directed by the present author and included W. Eller, R. Farr, L. Gleitman, W. Hall, M. Hoover, N. Mackworth, R. McDermott, R. Shuy, and T. Sticht.

ability to learn; (10) teacher's attitude: (a) toward the child's language, (b) toward the child's culture, (c) toward the child's ability to learn; (11) teacher's proficiency with the varieties of language, culture, cognitive style, personality, and ability present in the instructional setting, and with the instructional varieties present; (12) factors external to the classroom: (a) administrators' understandings and attitudes, (b) parents' understandings and attitudes, (c) community's understandings and attitudes.

It must be strongly emphasized that this set is not intended as an outline for a gigantic all-inclusive readiness test. Rather, it is intended to specify the kinds of variables that should be considered in developing and evaluating readiness assessments. The form and content of any particular readiness assessment should be determined by the instructional decision situation confronting the educator, and the specific kinds of reading tasks to be encountered by the student. Thus, once these goals are specified, a subset of factors could be selected from the overall list.

GUIDELINES FOR ADDRESSING THE QUESTION

Hoover (1974) and J. Mackworth (1974A) are against the use of tests that spring from current notions about reading readiness. J. Mackworth argues that "The whole concept of reading readiness is absurd. The child should be exposed to letters as early as possible in his life. If he is delayed a year or so he can only be the loser, whatever his potential learning ability may be....The only criterion should be the progress made during the year. If the child is unable to keep up with his peers, then he needs extra help." A number of points embedded in these comments deserve further consideration. The first sentence rejects the approach to reading readiness assessment that broadly states: "If the child demonstrates 'reading readiness,' then begin to teach him how to read; if the child does not demonstrate 'reading readiness,' then wait." This approach was discussed and dismissed earlier in this paper. J. Mackworth also states that delayed instruction is detrimental to the child's progress; here she sides with the position enunciated by Doman (1963).

In the last sentence in the excerpt quoted above, J. Mackworth remarks that if the child falls behind his peers, he needs extra help. This approach implies that readiness assessment is an ongoing process, trying to answer "What and how is the child ready to learn now in the area of reading skills?" Mackworth's comments lead to the following guideline for reading assessment: *Try to teach the child some reading skill; if the child has difficulty learning that skill, modify the task so that he can master it, or else try to teach him another reading skill that is easier for him to master.*

Hoover (1974) believes that "readiness tests should be eliminated. They are not particularly diagnostic and lend themselves to being used to 'sort' children." As seen earlier, a variety of evidence substantiates the view that currently available readiness tests do not provide useful diagnostic information. Hoover's criticism that tests are used to sort children raises several additional points. Sorting children into groups is harmful if these groups are inflexible. In order for a child to be placed in the instructional group that is most helpful to him, the appropriateness of these assignments must be checked frequently. Thus, assessment must be a continuing process. This is vital for two reasons. First, test scores can only provide estimates of reading performance, and this information is frequently undependable. Second, children learn at different rates; consequently, assigning a child to a particular instructional group at the beginning of a school year--or at the end of the preceding school year--should be viewed as a temporary arrangement. The instructional effectiveness of any "sorting" procedure should be re-examined after a short interval of time to assure the best possible match between the child's needs and the activities of different instructional groups.

Using the comments of Hoover and J. Mackworth as a point of departure, the following guidelines for the design and use of reading readiness tests are suggested.

Reading readiness assessment is a means of answering the question "What and how is the student ready to learn now in the area of reading skills?"

An effective approach to assessing readiness is to try to teach the student a particular reading skill. If the student experiences success in working with this task, then he is ready to learn it. If the student has difficulty, then the task should be reformulated so that the student is able to master it.

If the child's performance (either in learning to read, or on a reading readiness test) falls behind the performance of his peers, then he needs extra help.

Poor performance by a child on a reading readiness test is not a call to delay instruction, but rather a call to design and deliver instruction tailored to the needs of the child.

The instructional groups formed on the basis of a reading readiness assessment should be flexible, since children's learning rates differ, since reading readiness tests often provide undependable information, and since test scores are only estimates.

Reading readiness assessment needs to be a continuing process. The danger of a one-time assessment of reading readiness can be seen when Calfee, Chapman, and Venezky (1972) point to a study which showed that "the error rate on the second day of testing was about half that on the first day."

Reading readiness assessment should be integrated with instruction. Calfee (1974), for example, argues that reading readiness tests, as well as other kinds of school tests, need to be tied very closely to what is taught. Gates, Bond, and Russell (1939) studied four typical first-grade classrooms, using over 60 variables at the beginning of the year. These variables were then correlated with two reading achievement tests administered at three different times during that year. Since the correlations between reading readiness measures and later reading achievement measures were related to differences in instructional method, they concluded that "reading readiness tests, therefore, must be chosen to fit the teaching method. In other words one should test the reading abilities which the teaching program will attempt to develop in order to determine the needs of each pupil before instruction is begun and to predict the pupil's likelihood of becoming a successful reader." The principle that reading assessment should be tied closely to instruction implies that global or all-purpose measures of reading readiness assessment should be replaced by tests related to the specific reading tasks that will be encountered by the student. In order for this principle to be implemented, teacher training programs must be designed to help teachers become better informed consumers of reading research (and technology) (Wanat, 1973).

Assessment of a particular student's readiness for a particular learning task as well as evaluation of the appropriateness of the task for the student, can and perhaps should be carried out by the classroom teacher. The teacher is closest to the interaction between the learner and the learning task and therefore the best observer of the child's progress. This assumes, of course, that the teacher has had adequate training and experience in carefully observing students' learning processes.

It is useful to have a valid outside check on the teacher's appraisal of the student's potential.

Procedures for reading readiness assessment should make evident to the teacher the linguistic-cognitive strategies that the student employs in extracting meaning from written language. This is a reformulation of the "diagnostic" purpose of testing, and certainly this information cannot be provided yet by standardized tests of readiness. Nonetheless, it would be useful to have procedures for helping the teacher identify and attend to relevant linguistic- and cognitive-style differences manifested by the students. Also important are the

kinds of perceptual strategies that the student uses--how he applies his knowledge (e.g., his knowledge of language) to the task of making sense out of print (see, e.g., Wanat, 1968, 1972A, 1976A, 1976B, Wanat and Levin, 1968). Currently available reading readiness tests provide measures of the student's environemntal experience, since the items deal with the kinds of vocabulary, sentence structures, objects, and concepts to which the student has been exposed. Any reading test which confuses a student's readiness or reading ability with the student's familiarity with a particular set of test items provides the teacher with misleading information. A low score on this kind of test allegedly tells the teacher that the student has a low level of reading readiness or reading achievement. However, this may not be the case at all, since the student's low score may indicate a mismatch between his experiences and the vocabulary, syntax, objects, and concepts in the test. Hoover (1974) suggests that "tests should include vocabulary that not only gives all ethnic groups a chance but that gives students taught by different methodologies a chance."

Reading readiness tests should help teachers to observe and understand the student's learning processes in order to determine effective instructional strategies for those students. If a particular assessment procedure does not make a useful contribution to the educator's knowledge, why bother using it?

Reading readiness tests should directly relate test outcomes to instructional procedures. If a particular assessment procedure does not help the educator to select among instructional alternatives, why bother using it?

Reading readiness tests should be accompanied by appropriate interpretive scales. A set of test items alone does not make up a test that is meaningful to the teacher. A test also includes a set of scales that the teacher uses to interpret the student's scores. Therefore, the readiness test must be used in conjunction with interpretive scales appropriate for the students being tested. Thus, a number of different kinds of norms may be called for, each sensitive to different kinds of learner characteristics (language background, suburban versus urban versus rural, etc.). While the student's performance should be interpreted using norms based on the performance of students with the same characteristics, assessment of an educational system's overall success in improving reading readiness should be based upon a national or international standard. In this way, poor performance by a local educational unit can be identified, and inequities in the allocation of resources to different educational units, or to different populations of students, discerned.

CONSLUSION

Combining the set of guidelines in the preceding section and the descriptions of factors in reading readiness in the section before that, the following definition emerges: *"Reading readiness is the presence, in the learner and in the learning environment, of the skills, information, and attitudes that enable the learner to begin acquiring a new reading skill."* This definition emphasizes contextual factors--the characteristics of the learning environment--as well as the condition of the learner. This implies that any assessment of reading readiness must take into consideration variables such as the teaching methods and materials to be employed, characteristics of the teacher, and the attitudes of the learner's peer group toward reading. This type of assessment could in turn lead to a focus on modifying the learning environment and the nature of the reading skill to be acquired, as well as attempts to change the learner. Readiness, according to this definition is not a single, global entity, but can be assessed only in terms of specific reading skills.

Given this orientation, assessment of reading readiness does not relate solely to the child or adult who has not yet been taught to read, but becomes an appropriate educational procedure even for people who would already be considered mature readers. For example, one might ask if a particular college student is ready to read--or ready to learn to read--a particular psychology (or mathematics

or chemistry or law or medical) textbook. Or one might ask if a particular employee is ready to read--or ready to learn to read--a particular equipment repair manual, or a particular cooking and baking procedures manual. In order to answer these questions, one must be able to provide a detailed analysis of the task to be performed. While this point may seem obvious and trivial, commercially produced tests of reading readiness approach reading and the task of learning to read as a global and undifferentiated skill. An assessment of reading readiness can take place only after the assessor has specified the nature of the new reading skill that is to be acquired in sufficient detail so that prerequisite skills, information, and attitudes can be described. An assessment of reading readiness within this framework will pose the following question: *Do the learner and the learning environment display the skills, information, and attitudes that enable the learner to begin acquiring a new reading skill?*

Thus, the question "When is a child intelligent enough to read?" is a misleading formulation of an important problem. Trying to answer the question forces one to ask "What do you mean by *intelligent?*" and "What do you mean by *read?*" Attempts to identify a specific level of cognitive functioning as necessary for learning to read--such as a particular mental age--have not proved effective. The vagueness of the terms *intelligence* and *reading* provides little information for instruction and instructional decision-making. Rather than talking about "reading," we need to talk about *specific reading tasks.* And rather than trying to assess some general property of "reading readiness," we need to deal with *readiness to perform a specific reading task.* And finally, rather than ascribing all failures in learning to read to a student's lack of readiness, we need to begin to also examine contextual factors--characteristics of the learning environment--that might contribute to these failures, and to their remediation.

Consumer Awareness in Testing Reading

William Eller, SUNY at Buffalo
& Roger Farr, Indiana University

"Why is there any need for appraisal of children as they begin learning to read?" Once that question is answered, a second seems to follow: "What should be the nature of the appraisal for any purposes established in answer to the first question?"

The preceding paragraph states that the most important question the test consumer or the test developer needs to answer is "why is a test needed?" Of course, this does not mean that the test consumer's task is one of developing a rationale for each test he uses. Rather, it suggests a basic assumption about education. That is, that the educator is a decision maker, and that tests may provide information for making decisions. This assumption means that educators are aware of decision possibilities, that they are aware of alternative educational practices, and that they can delineate information that will help them make these decisions.

A second assumption which is basic to the framework presented in the latter part of this paper is that a test is a situation which has been structured so that certain behaviors can be observed and recorded. This very broad definition of testing is essential if we are to begin to consider the broad spectrum of information which is essential to educational decision-making. For example, when a first grade teacher talks to a kindergarten teacher to find out such things as the materials used for instruction during the previous year, how a given pupil responded to different types of instruction, or what interests a certain pupil manifested, that first grade teacher has, in fact, established a structured situation in which behaviors can be observed and information recorded. This is a test situation according to our broader definition. Of course, in the example above the observation is being done second-hand, but it nevertheless is a source of evaluative information on which instructional decisions can be based.

These first two assumptions--that tests should be used to provide information for decision-making and that a test is broadly defined as observation of behaviors under specified conditions--lead to a third assumption: that the makeup of a test is determined by the decision-situation to be served. This third assumption means that, while a test might be quite appropriate for one decision, it might be quite inappropriate for another decision. Each decision-situation establishes the parameters for the information needed. The decision will dictate whether information needs to be collected from every child or from a sample of children, how detailed the information needs to be, when the information is needed, and perhaps even the form in which the information will be most useful.

The fourth assumption on which the remainder of this paper is based is that the operational definition of reading or reading readiness will vary considerably from one teacher to another. Certainly teachers choose different skills to be taught to their children--e.g., the perceptual-motor proponents advocate the teaching of skills that are quite different from the skills advocated by the proponents of the language experience approach. This fourth assumption, when added to the other three, leads to the conclusion that different readiness appraisal materials (different "tests," broadly defined) should be available for the different operational definitions of reading readiness. Thus every teacher, whatever her perception of reading readiness, will be able to select evaluative materials which will provide information to guide the instructional decisions she must make.

The position that has been presented thus far is that testing is a means of providing information for making decisions. While this position may not seem unorthodox, it has rarely been followed in the U.S. schools. While millions of tests are administered to school children every year, very seldom are reasons or purposes stated for doing this. When there *is* a clearly-stated need for information, it is often the case that the test being used does not collect the information needed. How many times have we heard a teacher lament that the results of some test "don't tell me what I need to know." Finally, by developing a decision-making attitude, educators will increase their consideration of alternatives to present practices, and will increase their search for a greater variety of procedures for collecting the information they need for decision-maing.

The first step in operationalizing this view of testing is that one does not think first of "what test should I use?" Rather, one needs to first think about decisions. Put another way, the test consumer should not need to ask "What do the results of this test mean?" or "How should the results be used?" but rather should ask "Why was this information collected in the first place?" While this may seem like a minor difference, it is a shift in the test consumer's thinking from the position that "if some subskill is on a test it must be important," to the concept that "a test should provide specific information for making decisions."

After each decision has been stated, the decision-maker should list alternatives. Sometimes there are no alternatives and the educator finds he really has no decisions. Other times, attempting to delineate alternatives will cause one to search for additional alternatives. For example, in the case of reading readiness, this search might include attempts to develop different approaches to increasing children's language ability or different ways to organize the classroom for instruction.

Once the alternatives have been listed, the decision-maker needs to consider the information he needs to help him select the best alternative. Of course, the information needed is defined by the decision, by the decision alternatives, and by one's definition of the behaviors he is planning to make decisions about. It may be redundant to include the definition of the behaviors as an influence, since the definition will, in fact, have been a major influence on what decisions were listed in the first place.

When one considers the types of information needed for instructional decision-making, one should think not only of information about the student's behavior, but also about information regarding factors such as the teacher's behavior, the school environment, and parental attitudes. This should not lead to a wholesale listing of any kind of information one can think of. Schools are often guilty of collecting information that is not needed and that is never used. The criteria for determining the information that is needed center on the improvement of decision-making.

After completing the list of needed information, one should determine how the information is to be collected. The key criteria include the validity of the collection procedure (Is the information being collected the information that is needed?), the forming of the decision (Will the information be available when it is needed?), and the efficiency of collecting the information (Is the information worth the trouble it will take to collect it?).

In addition to considering the methods for collecting the information, thought should also be given to the subjects from whom the information needs to be collected. For example, for some decisions, information needs to be collected from every student, while in other cases only observations of random samples of students would be needed.

To briefly summarize, we have thus far established a point of view about the purpose for testing; we have put forth some basic assumptions regarding this point of view; and in the immediately preceding paragraphs, we have laid out a brief operationalization of the point of view. At this point, the area

of reading readiness and beginning reading achievement will be examined as it relates to the framework already developed.

In order to consider decisions, we need to know who makes the decisions. As far as beginning reading programs are concerned, classroom teachers, administrators, and parents are all significant decision-makers. The decision-matrix on page 31 includes several examples of the decisions, information they might need to make decisions, and some possible information-collection procedures. The decision examples shown in the matrix could be expanded or modified in many ways. They are merely attempts to emphasize the position that what is to be tested and how it is to be tested is dependent on the use to be made of the information. The operational definition of each of the items in the "information-needed" column will also influence what is to be tested.

The preceding paragraphs are summarized in the chart on page 32. It offers a series of contrasts between the traditional approaches to the appraisal of reading readiness and the more purpose-directed evaluation proposed. Certain characteristics of traditional readiness testing are listed opposite a description of a preferred alternative.

It should be pointed out that the approaches in the chart do not recommend the abandonment of standardized tests or of formal evaluation procedures. Instead, the emphasis is on appraisal oriented to the decision-making needs of the consumer, the educator. That sort of appraisal requires a much broader view of evaluation, as well as the integration of "testing," broadly defined, into instruction. Eventually, the phrase "diagnostic teaching" should disappear from the vernacular of educators, because appraisal has become so throughly woven into teaching that no one perceives of teaching without evaluation.

The goals and procedures proposed can only be achieved if those of us in teacher education and others of us in test-building and evaluation-planning can achieve an unusual degree of cooperation. Even then, it would take a while.

In the first part of this paper we have discussed the notion that any kind of test (however broadly "test" is defined) should be evaluated in terms of the purposes for which various educators (test consumers) need information. From this assumption, it is apparent that it is inappropriate to review a test as though it served (or failed to serve) the same purposes for all users. We will now consider criteria that test consumers should apply as they refer to the consumer information provided by formal test reviews.

Reviews of commercial tests, including reading readiness tests, tend to follow a traditional, if not prescribed, format. The customary layout begins with some vital facts about the test--facts useful to anyone who plans to purchase, or at least locate, some copies of it. Next, the review offers an overall description of major features--a listing of the subtests, a summary statement about the manual, enumeration of the scoring facilities, etc. Typically, there are observations about the content of the test, and some evaluation of the standardization, validity, and reliability (or at least of the publisher's procedures) in these three areas. Finally there is an estimate of the probable usefulness of the test for certain appraisal tasks. Many educators are so accustomed to this format for test reviews that it would not occur to them to question its appropriateness.

The next several pages are devoted to two distinctly different test reviews. The first review--of the Metropolitan Readiness Tests--is presented in the traditional fashion just described above. The second--a review of the Gates-MacGinitie Readiness Test--is offered in terms of certain purposes of different categories of test-users: teachers, administrators, consultants, paraprofessionals, etc.

CRITIQUE OF METROPOLITAN READINESS TESTS

The Metropolitan Readiness Tests represent about as orthodox a booklet of subtests as one could hope to find. Perhaps one indicator of the non-uniqueness

DECISION MATRIX FOR BEGINNING READING PROGRAMS

Decision Maker	*Decision*	*Alternatives*	*Information Needed*	*How Information Collected*
Kindergarten Teacher	I want to group my children according to overall reading readiness. I intend to have 3 groups. What groups should I assign each child to?	Each child can be assigned to 1 of 3 groups.	General reading readiness level of each child.	Continuous monitoring of each child's reading readiness by means of a checklist which operationally defines "overall reading readiness."
First Grade Teacher	I want to teach the children the reading readiness skills they need.	Each child be given some amount of instruction in each of the readiness skills.	Can any child already read (and therefore not need readiness skill instruction)?	A simple reading test.
			For those children who can't read, their ability to perform each of the readiness skills.	Continuous monitoring of each child's reading readiness skill development.
			What readiness skills did the children's previous teachers attempt to teach them?	Conversations with the children's previous teachers.
District Superintendent	Are the children doing adequately in readiness skill development? Are there any concerns in this area about which I might need to make decisions?	The superintendent can decide that everything is satisfactory and that the program can continue as it has, or he can decide to probe further if he decides a general problem exists.	The average reading readiness or beginning reading development of the children in the district. This needs to be compared to some absolute criterion level, or to some group that is similar	A single score general reading readiness test can be administered to a random sample of students in the district. Criterion- or norm-referenced interpretations can be made.
School Principal	What resources are needed to improve the reading readiness program?	A variety of resources can be added to the reading readiness programs.		
		In particular, these could include materials for teaching specific skill areas.	The average ability of the children in various skill areas.	A reading readiness test (single score) could be administered to a random sample of children.
		Teacher aides to assist teachers in whose rooms the most serious reading readiness problems exist.	Identification of children having the most difficulty learning reading readiness skills.	Each child could be administered a general single score reading readiness test.
		Field trips if students have limited experience backgrounds.	In general, what are the experience backgrounds of the children?	A questionnaire could be sent to the parents asking them to describe their children's experiential backgrounds. This could be followed by a teacher interview with each child.
Board of Education (Parents)	Are we getting our money's worth out of our reading readiness program?	Lessening or increasing the support for the program—both psychological and material.	How well do children learn after they have gone through the readiness program?	Reading test scores of children in the first grade compared to children who come from similar backgrounds, but who have not been exposed to a reading readiness program. For this test, only a random sample of children would need to be tested on a test of general reading ability.
			Do children engage in reading?	Surveys of the reading habits of a random sample of children in the district.
			Do children increase their reading readiness after being in the program?	A general reading readiness test (single score) administered to a random sample of children in kindergarten and beginning first grade.

APPROACHES TO THE APPRAISAL OF READING READINESS

Traditional	*Proposed*
A commercial readiness test is selected for non-decisions making reasons, such as previous use in the school, administrative convenience, familiarity to examiners, cost, relationship to other materials from same publisher, etc. In this situation, the test exerts undue influence on the instruction.	The appraisal opportunities and materials are determined by the prevailing program and philosophy demanded by the decision situation and by the available alternatives. The curriculum governs the evaluative processes, which in fact, *are a part of that curriculum.*
Appraisal of readiness is not continuous; perhaps it is even a one-time occurrence.	Readiness evaluation is continuous because it is a part of instruction. Ideally, the evaluation should be so thoroughly integrated into instruction that it is not viewed as a discrete process.
Readiness appraisal is heavily dependent on input from one or more commercial tests, with some tempering effect based on teacher observation and informal testing.	Readiness appraisal is not predictably dependent on any single type of evaluative procedure(s). Instead, the teacher and other staff draw from any available sources of decision-making information pertinent to present alternatives.
Evaluative processes focus upon appraisal of the pupil and his characteristics, with occasional attention to the curriculum and its materials.	Appraisal of readiness includes process evaluation as teacher is led to examine his own procedures and methods.
The test-builder is heavily product-oriented; that is, his test specifications are determined mainly by his perception of the skills manifested by pupils.	The test-builder is at least somewhat process-oriented, and the test reflects awareness of the educator's procedures in reading the necessary decisions.
Concurrent validity has considerable importance.	Concurrent validity, in terms of correlations with other tests, is less important. Content validity takes on greater significance.

of the Metropolitan is its adoption, with only a change of title, by the New York State Department of Education, an agency not famous for innovation in measurement. Because it is utterly typical of the readiness tests of the past 25 years, it does not rate favorably on some of the criteria previously discussed. It will be evaluated in terms of those criteria several paragraphs hence. Preceding that is a review of the Metropolitan as the traditional test that its authors evidently intend.

General Description of the Metropolitan Readiness Tests

The current version of the Metropolitan was copyrighted in 1966, but there have been several previous editions, the earliest of which appeared in 1933. The six main subtests are packaged together in either of the two forms, A and B, and the total administration time is 60 minutes. These tests are entitled Word Meaning, Listening, Matching, Alphabet, Numbers, and Copying. There is also an optional seventh test which is a modification of a part of the Goodenough-Harris Drawing Test of 1963. The manual of directions recommends the administration of the first six tests in pairs at three separate sessions, and the use of a fourth session if the seventh test is given.

The Metropolitan Readiness Tests are perceived by the authors as indices of general readiness for learning in first grade, rather than focused on readiness for reading. The physical, statistical, and editorial aspects of the Metropolitan Readiness Tests and the manual are representative of the high standards of the test department at Harcourt. In the manual, the directions to the examiner are lucid and easy to follow. The directions which the examiner is to give orally to the pupils should pose few comprehension or execution difficulties for children in kindergarten or first grade. Color, type-sizes, and spacing have been used to guide both test administrators and examinees.

Of the 36 pages in the manual of directions, a good 20 pages are devoted to statistical data on norms, validity, reliability, and standardization. From the standpoint of the measurement specialist, a lot of useful information is provided. But there is the change that domination of the booklet by so many pages of data may "turn off" some classroom users of the Metropolitan. Perhaps there is some way in which these kinds of data could be presented for those who need them without giving the manual the appearance of a report from the Bureau of the Census.

Content of the Subtests

The individual subtests in the Metropolitan Readiness Tests booklet are relatively short (as few as 14 items), and some are extremely diverse. Thus, they are not well-suited for use in diagnosing areas of underdevelopment. However, in the manual, the authors or publishers have wisely pointed out the inappropriateness of attaching significance to the separate scores from the subtests. Certain users have felt that at least some of the subtests are not named very precisely. The Matching test, for example, would be described as a test of visual discrimination by many users. It includes items requiring discrimination of pictures and geometric forms, as well as of words and letter combinations. Considering the work of Muehl and others in the early 1960s, it is surprising that these nonverbal items have not been replaced.

The Metropolitan differs from many other readiness tests in that no subtest of auditory discrimination is included. Auditory functions are extremely important in the total test, but they involve listening comprehension rather than discrimination of sounds, even though the authors list auditory discrimination as one of "the important components of first grade readiness." Two other aspects of readiness not isolated by the Metropolitan are (a) ability to follow oral directions and (b) ability to identify letters, words, non-words, and numbers. These skills are important in the appraisal of readiness because until the ex-

aminer knows that pupils have sufficient ability in following directions and in identifying certain kinds of symbols, he cannot be sure that they know what they are being asked to do on other subtests. In a sense, these two abilities--and perhaps certain others--are a part of the readiness to take a readiness test.

Validity

About a dozen pages of the administrator's manual are devoted to setting forth the company's case for the validity of the Metropolitan Readiness Tests. As might be expected, the case is made in terms of (a) content validity--the extent to which the tests deal with the areas deemed important by experts or research; (b) concurrent validity, in which the results are compared with other established measures; and (c) predictive validity--the extent to which the tests predict school success or non-success for first grade children.

In the area of content validity, the manual provides extensive descriptions of the functions and purposes of the several subtests. But because this material is, of necessity, more subjective than the other two treatments of validity, there is greater need for the test user to strive to understand the precise intent of the authors. Also, there is a considerable prospect that the user will disagree with the authors, either because he does not understand their position, or because he is convinced that they have not accurately described one or more of their subtests. So, on the one hand, the content validity material in the manual is the most useful to the prospective examiner, and on the other hand, it requires the most careful consideration.

Concurrent validity is established for this edition of the Metropolitan by correlating scores with scores from the Murphy-Durrell Reading Readiness Analysis and the Pintner-Cunningham Primary Mental Ability Test. Less detailed comparisons are also made with three other measures of mental maturity, as well as with the Goodenough-Harris mentioned earlier as an optional seventh subtest. The various correlations between the Metropolitan and these other tests fall in the ranges expected by anyone who has studied a number of test manuals over the years. For whatever it is worth, the Metropolitan correlates fairly highly with the Durrell and the Pintner-Cunningham.

The predictive validity data also provide no surprises since the correlations between September scores on the Metropolitan and end-of-first grade scores on the Stanford Achievement Test fell between .57 and .67.

Summary

This sort of "mental-measurements-yearbook-type" review probably only deserves a one-sentence summary: "The Metropolitan is a typical readiness test in content and function, and is better than average in those details provided by the publishers." To put it a bit more strongly, only the first sentence of this review was necessary. A lengthy review of the Metropolitan or another typical readiness test is not useful to teachers, consultants, administrators, or parents because most of the information does not help them to make the decisions they need to make regarding youngsters who are approaching early reading instruction. In fact, the information does not enable professional educators to make a genuine choice between these tests, even if they have already decided to measure readiness in the classic fashion. The information provided under the heading "content validity" in the manual might seem to help the test-user at least to say to himself: "This test is exploring the skills and knowledges that I am certain are important." But the concurrent validity correlations hovering around .80 suggest that other tests must be measuring many of the same basic functions, and would thus be quite suitable also.

An examination of the widely used readiness tests, accompanied by study of the concurrent validity data and the predictive validity correlations, leads to the notion that most of them are designed to measure some kind of general verbal

ability. Since the subtests are, as a rule, admittedly not long enough to be reliable, differences in subskills measured are not dependable bases for selecting one test over another.

Even more unfortunately, the Buros-type reviews of readiness tests tend to perpetuate the status quo of readiness appraisal. They reinforce the notion, for example, that a one-time measurement of readings is permissible, when many aspects of readiness need continuing appraisal. They strengthen the tendency to ask what Durkin has referred to as a "wrong question:" "Is the child ready to begin learning to read?" and add to the accompanying indifference to the methods and materials which are likely to be used with the child. They add to the already excessive respect for predictive validity correlations, as they provide teachers with a handy excuse for the limited achievement of pupils who earlier had low readiness scores. And because they are so scientific-sounding, they make the formal readiness test appear more important than it really is in the total process of evaluating readiness. But the most serious problem with the orthodox reviews of reading readiness tests is that they do not evaluate the tests in terms of the larger perspective of the evaluation of pupils' readiness as already discussed. The next review of the Gates-MacGinitie Readiness Test will be directed toward this broader perspective.

GATES-MACGINITIE READING TEST: READINESS SKILLS

As was stated, it is only appropriate to evaluate a particular test in terms of why the test is to be administered. Therefore, this review of the Gates-MacGinitie Reading Test: Readiness Skills will be related to two specific reasons why the information divulged by the test might be needed.

One reason that a reading readiness test might be administered by a kindergarten or first grade teacher is to determine which reading readiness skills need to be further developed by the children in her class. The decision-matrix for this decision could include the following:

DEVELOPING READING READINESS SKILLS

Decision	*Alternatives*	*Information Needed*	*When Needed*
What reading readiness skills should I teach to my children?	Varying amounts of instruction in skills which the children have not mastered and that they need to master. This list would include visual discrimination, auditory discrimination, language development, ability to follow directions, ability to learn to recognize words.	How well has each child mastered the skills listed?	Continuously with instruction.
		Can any of the children already read?	When instruction is being planned for the first time.
		Each child's attention span and readiness to attend to the kind of instruction that will be provided.	Continuously with instruction.
		Each child's background of experiences.	When instruction is being planned for the first time.

The information is this decision-matrix is based on a specific definition of the

area to be tested. That is, the definition of the skills necessary for beginning reading are explicity stated. This is in keeping with the position taken in the first part of the paper that each test consumer must continuously study this complex process we call reading if he is to make conscious decisions about what to teach and how to assess those behaviors. The alternatives column includes some areas that several people would not include and omits others that some think should definitely be included.

In addition to stating his needs to use a test (that is, to collect information) and what information he needs to collect, it is essential that the test consumer be able to verbalize his operational definitions of the skills. For example, what does one mean by "auditory discrimination" or "language development," or "the ability to read?"

Perhaps, at this point, one might suggest that it is beyond the teacher's ability or perhaps even responsibility to engage in this task of stating decisions and defining the information needed for making decisions. One might argue that it is only the test author's job to define what is to be tested. However, suppose that Siegfried Englemann and Roach Van Allen each decided to build a reading readiness test. It is obvious to anyone who has followed the work of both of these men that they would each include quite different skills in their tests. Those differences in test content and format would be based on legitimate differences in their conceptualization of what a child needs to learn in order to become a reader. The teacher must either blindly accept one of these operational definitions, or she must consciously sort out in her own mind what she believes the process to be and select a test accordingly. The latter alternative is preferable, especially if one agrees that a teacher should have the process of reading (or reading readiness) clearly in mind as she selects instructional materials and procedures throughout a school year.

The teacher's decision-making responsibility is an unavoidable reality. Either the teacher decides to make decisions consciously, or decides to ignore alternatives to present instructional practices. In the latter case, the teacher becomes less than a teacher and could probably be replaced by a teacher aide or a mechanical teaching device. It is the teacher's thinking, i.e. decision-making ability that makes her indispensable in the classroom.

Now that the basic arguments have been presented that it is essential to state decisions and to define the information needed for those decisions, the first example can be expanded and we can evaluate the Gates-MacGinitie test in relation to the stated decision. See the decision-matrix on page 37. It should be re-emphasized at this point that the critique of the Gates-MacGinitie test is based on a particular purpose for which the information is needed and, on a particular definition of reading readiness. If the purpose for which information is needed changes, of if a different set of skills are defined as being part of reading readiness, then a different base for evaluating the test would be established. Certainly, this means that the single test review which can serve all test purposes is of extremely limited use.

In summary, it appears that the Gates-MacGinitie Readiness Test falls far short of being useful to the teacher for continuous instructional decision-making. Any teacher who has worked with children knows that children often develop the skills that are being taught quickly. If a test is administered one week and not scored and returned to the teacher until the next week, it is quite possible that the test is no longer a true representation of the readiness skills of the children in the class.

A second example of an instructional decision can easily demonstrate different purposes require different tests. Suppose that the superintendent of a large school district has decided to utilize additional resources (teacher aides, money for instructional materials, smaller class sizes, etc.) in the schools in his district where students have the least-well-developed reading readiness skills. The decision matrix might look like the one on page 38.

GATES-MACGINITIE DECISION MATRIX

Information Needed (Including Operational Definitions)	*When Needed*	*How Well Does the Gates-MacGinitie Test Meet this Need?*
Each child's reading readiness skill ability:		
a. Visual Discrimination. The ability to see likenesses and differences in words and letters when those are presented in isolation and in context.	Continuously as instruction is being planned.	The visual discrimination test only provides for a single assessment, so the test is not useful for continuous instructional planning. The discriminations are with words only and are presented in a context that is unrealistic to the reading task.
b. Auditory Discrimination. The ability to discriminate between sounds that make a difference in the meaning of words.	Continuously as instruction is being planned.	The auditory discrimination test only provides for a single assessment, so the test is not useful for continuous instructional planning. The test is one in which the examinee must retain the definition of a vocabulary item, and then match a spoken word with the picture. This test is a different operational definition than the information needed.
c. Language Development. The ability to form oral language that communicates in terms of the typical classroom language. That is, the child's ability to speak in understandable semantic and syntactic patterns.	Continuously as instruction is being planned.	Not included on this test.
d. Ability to Follow Directions. The child's performance in understanding direction words that are used in a beginning reading program.	Before the administration of a test that uses these directions, and continuously as this instruction is being planned.	The "following directions" test only provides for a single assessment, so the test is not useful for continuous instructional planning. The test should be administered first in the test battery, and should include such basic things as whether the children can follow a direction such as "put an *X* on," or whether they know what is meant by *row*. If a child fails such a test, he should be taught these things before the rest of the test is administered. The Gates-MacGinitie test assesses such things as whether the child can select the shortest word, the tallest flower, the farthest away. These are concepts that are useful to a child, but they are not the instructional directions used in a beginning reading program, nor are they the "following directions" concepts a child needs to perform on this test.
Ability to read (Can any of the children already read?): A test of whether a child may have learned to read and therefore has no need of instruction in readiness skills. The test need not provide a detailed analysis of reading ability.	Before a readiness test is administered.	The Word Recognition test is given at the end of the readiness battery. It should be given first, and those that perform well on the test (14 correct or higher according to the test manual) should not be given the remainder of the readiness test. The actual task performed by the child (matching a word the teacher reads with a written word) is not so closely related to reading as having the child produce the spoken representation of a word he sees in print.
Each child's attention span and readiness to attend to the kind of instruction that will be provided: This is defined as a child's response to the reading instruction he will be exposed to. He should be exposed to samples of this instruction and his behaviors observed.	Continuously as instruction is being planned.	While it is suggested in the test manual that this information be collected, no procedures or provisions for collecting the information are provided.
Each child's background of experiences: This is defined primarily as his oral vocabulary or concept development. The vocabulary or concepts should be those he is apt to encounter in his beginning instructional program.	When instruction is being planned for the first time.	While it is suggested in the test manual that this information be collected, no procedures or provisions for collecting the information are provided.

SUPERINTENDENT'S DECISION-MATRIX

Decision	*Alternatives*	*Information Needed*	*When Needed*
To which schools should extra resources be allocated?	Any of the schools in the district.	The average reading readiness development of the children in each school.	At the time the decision about resources is being made.

Information Needed in Operational Terms

The administrator does not need a thorough analysis of each child's reading readiness skill development for this decision. Therefore, it is not essential that he define the skills to be taught in each classroom. In fact, it is not even essential that each child in each school be tested; it would be quite possible to randomly sample children from each school and test a much smaller population. Because the superintendent is interested only in the general level of reading readiness skill development, and because of the very high intercorrelations between all of the subtests of reading readiness, he would only need to administer a test which contained one subtest. That subtest should be the one which correlates highest with all of the other reading readiness subtests.

Therefore, the superintendent might find that the Word Recognition subtest, which is an excellent predictor of reading achievement and which correlates highly with all other reading readiness subtests, might be the best test to use as a means of gathering information about the average reading readiness skill development of the children in each school. This test might be 25 to 30 items long instead of only the 14 included in the Gates-MacGinitie test. The added length would increase the reliability of the assessment of each individual, and thus the reliability on the average for each school. Another possibility is for the administrator to select two subtests from the Gates-MacGinitie test and combine these into a single test. Because of the high correlations between the various subtests, adding a second subtest is as appropriate as lengthening the first test.

If such a test--a 30-item test of word recognition--were constructed, and if it were administered randomly to every fourth child in each school in the district, the superintendent would have the information he needed to make his decision.

From comparing the preceding two decision-situations, it is obvious that the test that would be appropriate for the administrator's decision would not be appropriate for the teacher's decision. Without considering *why* we need information and *what* specific information we need for making decisions, we cannot be discriminating test consumers.

In order to complete this review, it seems that a more complete analysis of the subtests should be made. This further review is one which is based on the present writers' overall definition of reading readiness, and on the further operational definitions of each of the specific sub-areas. In addition, this review is based on the notion that the test is to be used for making instructional decisions about skills to be taught.

Listening Comprehension. According to the publisher, this test "measures the child's ability to understand the total thought of a simple story." Many of the questions seem to measure retention of factual details rather than "understanding the total thought." It would have been very useful if the publisher had provided an outline of the comprehension skills to be assessed by the 20 items.

Auditory Discrimination. According to the publisher, this "subtest measures the child's skill in distinguishing between two words of similar sound." The test seems to call for very careful listening on the part of the child to see if

he can distinguish between word pairs like *cup* and *cub*. Obviously, a child who has learned to talk and who can understand what he is to get when he is told to go to the kitchen and get a pail or a pan has been able to discriminate the phonemes of his language, or he would not be able to respond to the directions and, furthermore, he would not be able to talk.

Visual Discrimination. According to the publisher, this test "measures the child's ability to distinguish between the printed forms of two words." In reviewing this test, one wonders if the child knows what he is asked to respond to. For all children who cannot yet read, the differences between word pairs like *beet* and *beed* are differences that do not make a difference. Certainly, the child should be taught what he is to respond to before he is asked to perform the task. Perhaps a better task is to determine if a child knows what a word is, or what a letter is, or whether he, in fact, knows what reading is. These are visual discrimination subtasks that are vital to beginning reading.

Following Directions. According to the publisher, this test "measures the child's skill in following increasingly more complex directions." The first six of the 14 items in this test measure the following:

Test Item	*What Is Actually Being Assessed*
1. Mark what the girl is looking for.	Testing comprehension.
2. Mark the one that is the same size as the middle one.	Visual discrimination.
3. Mark the tall girl wearing a hat and the little girl who does not have any hat.	Visual discrimination of size.
4. Mark the ducks that are farthest away.	Visual discrimination of distance.
5. Mark the boy who is closer to the door and mark the girl who is standing on the other side of the door.	Listening comprehension and interpretation of pictures.
6. Mark the first and last hats in the row.	Understanding of the direction-words *first* and *last*.

These test items seem to be assessing something other than the "following directions" concepts that are needed by the beginning reader. Item six seems to come closest, but it is quite confusing because it is an item requiring the child to make two things. If the child puts an *X* on the first two items, do we assume he knows *first*, but that he does not know *last*?

Letter Recognition. According to the publisher, this test "is designed to measure the child's recognition of letters of the alphabet." While this skill is highly predictive of reading ability, teaching the names of the letters of the alphabet has not been shown to affect reading ability. If this is true, the behavior does not need to be taught or, of course, assessed. This raises the question of the diagnostic validity versus the predictive validity of reading readiness tests. Merely because one thing is predictive of (correlated with) a second thing, it does not mean that teaching the first thing will cause the second thing to happen.

Visual-Motor Coordination. According to the publisher, this test "measures the child's skill in completing printed letters." Probably this test was included in deference to the perceptual-motor skill devotees. According to the

definition used in this analysis, this is not a skill to be taught or tested as part of reading readiness.

Auditory Blending. According to the publisher, this test "provides information about the child's ability to join the parts of a word, presented orally, into a whole word." According to the definition of reading readiness used in this analysis, this is not a skill to be taught. No one learns to pronounce a word by blending the parts that he has listened to into a whole. The child is almost always told a whole word and he mimics that whole word, usually in a meaningful context. Only when a child has extreme difficulty with a word such as *spaghetti* do parents or teachers try to break a word into parts. In any event, this task seems to be a weak attempt on the authors' part to get at a word-attack skill, that of blending, which itself has been criticized by linguists as being an unreal reading task.

Word Recognition. This test has been dealt with elsewhere and will not be reviewed further at this point.

In all fairness, the Gates-MacGinitie is a fairly good example of the kinds of reading readiness tests that have been in use in America for the past 50 years. It is not the technical expertise of the test publisher or the test authors which is, in fact, being criticized in this review. Rather, what is being criticized is the whole concept of testing.

We should gradually eradicate the concept of the all-purpose test; these tests should be replaced with specific tests for specific purposes. Secondly, we need to develop techniques of assessment that are inexorably intertwined with instruction. Third, we need to develop teachers who are conscious of their beliefs about the skills of reading and who use those beliefs to select tests. These efforts will demand much of all of us, but the current concern about the value and use of tests, and concern about the monumental waste of energy and money in giving tests that provide little useful information for instructional decision-making, should cause us to become more conscious of the need for increased consumer awareness about test reviews, and more generally, of the need for increased consumer awareness in testing reading.

What Should Early Reading Tests Measure?

J. Japp Tuinman,
Simon Fraser University

Between 150 and 250 million tests are administered annually to public school students (Kirkland, 1971). That total represents three to five standardized tests per pupil. In view of this high figure, an important question is: How early do we need "early reading tests?" The most important consideration in trying to answer that question is the instructional utility of the information yielded by the test.

I see no benefits in instructional decisions based upon general prediction instruments such as IQ tests or independent readiness batteries. To be informed that a particular student most likely will be a low achiever has very little instructional value. As to reading tests, the fact that such batteries are independent of a specific implemented program of instruction makes them, in many instances, less than optimally relevant. MacGinitie (1969) makes a sensible distinction between the questions: "Is this child ready (often researched)?" and "What and how is this child ready to learn (seldom researched)?"

The classroom teacher is just about the only one who knows what is expected of the child. Since our lack of understanding of "basic" predictors limits us to "real-life" tasks, she is in a much better position to select readiness probes than is any test author. As a matter of fact, I think that it is extremely important for us to train teachers how *not* to use tests.

If the teacher is inexperienced or inept, and she utilizes a teaching program that is very precisely described, it might be desirable to supply her with test items. In addition to possessing sufficient technical quality, these items should have a degree of content validity unusual for most ready-made readiness or achievement tests.

In reality, such program-specific test items double as readiness and achievement measures. The degree of their adequacy in the former sense depends on the similarity between the item and subsequent learning tasks.

To many measurement experts, the thought of turning a teacher loose with a set of items that have not been prepackaged in a test with known psychometric properties, is horrifying. Are we returning to primitive behaviors long outgrown? The issue of balancing the technical adequacy of the measurement and its instructional utility is a complex one, but let me make a few observations. First, we may have pushed technical excellence beyond the level where further increases are relevant. Some recent research shows that not all violations of sacred dogmas of item writing have devastating effects on the validity and reliability of the measurement (Baker, 1971; Board and Whitney, 1972; Marso, 1970; McMorris, Brown, Snyder, and Pruzek, 1972). Obviously, this does not mean that we can abandon care in item constructions. It does mean that "purity" is not as severe a constraint as is sometimes thought. More important, however, is a second observation: It makes no sense to use a test that provides more precise information than can be used effectively for making instructional decisions. In the practicality of classroom instruction, all that is often needed are measures that relegate a student to one of relatively few categories. For such measures, standard errors of measurement can be tolerated that are inexcusable when associated with a regular standardized test. We should be wary of overextending the criteria that are desirable in the latter case to measures that have an entirely different function.

In the preceding paragraphs, I have made some remarks about classroom measurement. I indicated that the only "readiness" assessments that are accept-

able to me are teacher judgments based on close observation of performances in everyday tasks, and the administration of test items that parallel the actual instruction to which the child is exposed. Standardized tests of general mental abilities, tests of readiness, and tests of achievement at the end of the first year, have little merit in beginning reading instruction.

Let me now turn in a more positive direction. One thing that is needed in the arena of measurement at the early stages of reading are tests based on an explicit model of acquisition of reading skills. In recent years, an enormous amount of literature on "reading models" has been collected and analyzed (see Davis, 1971; Singer and Ruddell, 1970, 1976). While it may be true that many of these "models" are not really models, and while others may have received only scant empirical verification, there is, in fact, some convergence among the models with respect to common components. We should emphasize the development of tests and teaching sequences on the basis of these models. To the extent that current readiness tests are based on any theory at all, they are usually based upon a very broad model of intellectual development. It would be very beneficial if we became more specific in identifying, measuring, and (possibly) modifying the processes underlying learning to read. For practical examples of what I mean, one can refer to the work of Calfee. His Basic Skills Test consists of a large number of subtests, justified on the basis of his three-stage acquisition model (see Calfee, Chapman and Venezky, 1972). Among the subtests are: auditory matching, segmentation, correspondence learning, sequencing, and memory for visual forms. In all, five areas of cognitive functioning are tested: matching of visual forms, letter-sound association, auditory-phonetic identification, vocabulary knowledge, and general achievement.

I am not specifically arguing for the validity of the model or measures proposed by the Calfee group. I am merely saying that systematic development of early reading tests, on the basis of some explicit model of reading acquisition, is needed more than any new twist in the old array of subtests that are now common fare in readiness assessment.

The eventual outgrowth of such efforts will be tests much like the ITPA, but related to reading, rather than to the communication process in general. In addition, the relationship between subtests and actual reading behavior most likely will be more detailed and more explicitly defined than is the case for the relationships between the communication process and the ITPA subtests.

A final observation: Rohwer's (1971) work on the measurement of intellectual abilities of young, deprived children has some definite implications for any assessment of readiness we might want to engage in. Rather than assessing whether or not a child has learned something, it may be valuable to find out whether he *can* learn it if we just sit down and try to teach it to him. The contrast is somewhat overstated, but how important is a lack of readiness that can be erased in a half-hour teaching session? How strongly do we want to argue that this lack of readiness signals lack of ability rather than lack of opportunity?

IQ Is and Is not Related to Reading

Harry Singer,
University of California, Riverside

My apparently paradoxical title emphasizes the variable relationship between IQ and reading. The explanation for this variable relationship is that if a particular reading task (such as acquisition of symbol-sound correspondence) is within the mental age range of a group of students and if instructional conditions allow adequate time for achieving the task, then IQ may have a significant relationship to *rate of acquisition*, but not to *accomplishment of the task*. However, if the reading task challenges even the most capable students in the class and time or trials for learning the task are limited, then IQ is likely to be highly correlated with achievement of the task.

This explanation is consistent with Cook's (1951) generalizations based upon his review of research literature on learning and individual differences. While acknowledging difficulties in measurement, Cook nevertheless concluded: "...if the responses to be learned are sufficiently simple and the goals that have been set so limited that a high proportion of the group can master them during the period of learning, the variability of the group becomes less; if the task is complex and the goals unlimited, so that the abilities of the most apt members of the group are taxed during the period of learning, the variability of the group increases." Simple tasks are those in which the variability of a group decreases as a function of practice. Complex tasks are those in which the variability of a group increases as a result of instruction and cannot be mastered by the group, regardless of the time devoted to them (Bloom, 1971; Cronbach, 1971; Peterson and Barlow, 1928).

Simple and complex tasks in reading exhibit these variabilities. For example, variability in acquisition of word recognition abilities (such as symbol-sound correspondence) *decreases* at successively higher grade levels. On the other hand, the range in achievement in word meaning increases throughout the grades. So does the range in mental age and the variability in the IQ's of bright versus average versus dull (Bayley, 1946, 1955; Cook, 1951). Hence for members of a particular group, the correlations between IQ and word recognition abilities such as symbol-sound correspondence *decrease* while the correlation between IQ and reading comprehension increases.

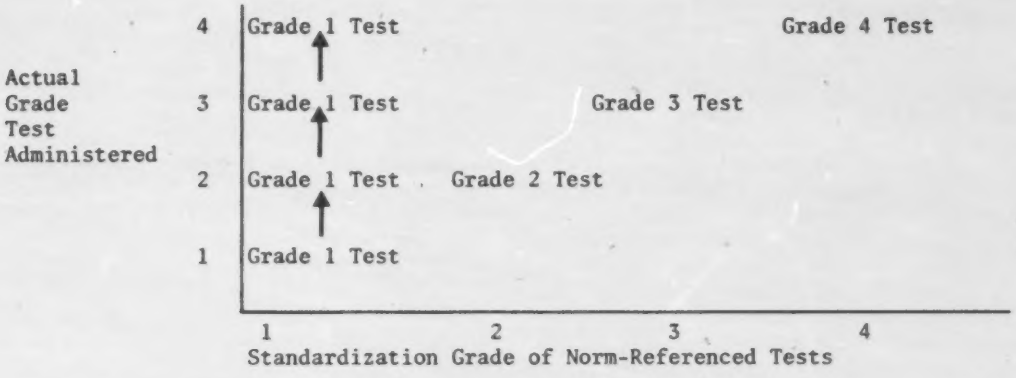
The term "reading" is, in fact, used ambiguously to refer to reading acquisition or to reading comprehension. The ambiguity is further compounded when the developmental stage of the reader, which can vary from beginning to skilled reading, is unspecified (Wiener and Cromer, 1967). Beginning readers still in the process of acquiring implicit rules for relating orthographic symbols to linguistic forms are quite variable in this process of reading. Hence, the relationship between IQ and these components of reading would still be high at this stage of their reading development.

However, skilled readers are not only less variable in word recognition abilities, such as symbol-to-sound relationships, but they also seem to have short-circuited them. Instead, they engage in "lexical reading," which tends to bypass phonological processes and involves a more direct correspondence between graphemic cues and lexical meanings. At this stage of reading development, individuals may perceive words as though they were ideographs (Smith, 1972). Apparently, in the highly skilled reader, lexical meanings are anticipated, triggered, and confirmed by perceptual sampling of graphemic cues (Hochberg and Brooks, 1970). For skilled readers, then, there would be little or no variability in accuracy in word recognition ability. Hence, the correlation between their

IQ and reading acquisition behavior would be zero.

We can demonstrate these changes in variability and resulting changes in their correlations with IQ for members of a particular age group. For example, as shown in Figure 1 below, if we administer equivalent forms of the *same* norm-referenced *reading achievement* test given at the end of grades one, two, three, and four to each successive grade level, the correlation between IQ and performance on this test will steadily drop as the group progresses through the grades. The reason for this decreasing trend is that we will have kept the reading task constant while members of the group were learning to read and developing in abilities requisite to performance on the reading test. As they did so, their mean performance on the test would gradually approach the top score and their variability on the test would be reduced. In short, they would tend to have learned to read and would therefore have mastered the test. As they did so, the *correlation* between performance on this test of reading acquisition and performance on the IQ test would decrease.

FIGURE 1



*Schematic Design To Show Norm-referenced Tests for Assessing Reading Acquisition and Norm-referenced Reading Achievement.**

This use of the norm-referenced grade one reading tests contrasts with the usual way in which tests are constructed and administered. Norm-referenced tests are designed so that the average individual at each grade level gets only 50 percent of the items correct on the test. This test construction objective is obtained in reading tests by using more difficult words, and by using longer, more complex sentences on successive grade level tests. Thus, the typical norm-referenced test administered in the elementary grades assesses *not only* development

**The diagram shows that the grade one test is also used in grades two, three, and four as a type of criterion-referenced test. The first column shows the grade one test administered at the end of grades one, two, three, and four for those students who have not yet mastered the test. The diagonal norm-referenced tests are administered as usual at the end of grades one, two, three, and four to all children in these grades.*

The column one norm-referenced test administered in grade one can also be given as a criterion-referenced test to assess reading acquisition *over grades one, two, three, and four. The diagonal tests provide the usual norm-referenced test information on relative reading achievement. Under these test conditions, correlations between IQ and criterion-referenced tests over grades one through four (column one) decrease and between IQ and norm-referenced tests over grades one through four (diagonal) increase.*

of reading acquisition behavior but also achievement in knowledge of morphemes, vocabulary, and other cognitive abilities tapped by the longer and more complex sentences.

Essentially, the norm-referenced test increases in difficulty because it gradually shifts from measurement of reading acquisition behavior to measure of cognitive and lexical development--the very same tasks usually included in tests of intelligence. Under these test conditions, the correlation between IQ and reading, of course, remains at a high level for members of a group, even though they had learned to read. Also, under these test conditions, the teacher has only one test score which cannot be separated into reading acquisition behavior and general reading achievement.

The remedy is quite clear: Separate the definitions and assessment of *reading acquisition behavior* from *reading achievement*. One way this separation can be achieved is to administer in grades one, two, three, and perhaps four, either the same or equivalent forms of a given standardized, norm-referenced test that is usually given only at the end of grade one. This annual, repeated testing would, of course, only be administered to those students who have not yet mastered the test.

Used in this way, the *norm-referenced* test would act like a *criterion-referenced* test because all the components of the test would be kept constant while students were learning to read and improving in reading achievement. Keeping the task constant is a necessary condition for assessing any kind of learning. Thus, progress in learning to read could be readily measured and separated in the primary grades from general reading achievement.

Reading achievement, as traditionally measured by norm-referenced tests, can co-occur with assessment of reading acquisition behavior at grades two and above by administering the usual norm-referenced tests given at these grade levels. From these two types of assessment, teachers would then gain diagnostic information about an individual's progress in learning to read in relation to first grade norms and his general reading ability in relation to his grade level or peer group.

Norm-referenced tests given at the ends of grades two, three, and four and higher grade levels can also be repeated at successive grade levels to measure reading acquisition behavior on more complex words and sentences.

Although my paradoxical title has now been explained, further clarification and particularly evidence to support my position still needs to be adduced. For this purpose, I shall briefly review the nature and development of both reading and IQ, then draw upon theoretical and empirical evidence on conditions of schooling and instruction which influence the relationship between IQ and reading.

NATURE AND DEVELOPMENT OF READING

An individual formally starts to learn to read when he is at a stage where he can mobilize almost fully-developed linguistic abilities (Ruddell, 1970; Singer, 1972). As early as the first grade level, he can utilize his phonological, morphological, syntactic, and lexical abilities for responding to print. He is likely to do so, if instruction is based on connected discourse, instead of isolated letters or words. For example, an analysis of errors in oral reading or connected discourse in the first grade indicated that the errors were predictable from the syntactic and semantic constraints in the parts of the sentences preceding the error (Weber, 1970). Furthermore, Goodman (1965) has demonstrated that contextual constraints facilitate and augment accuracy in word recognition during the process of reading.

As an individual progresses in reading, he first utilizes linguistic constraints for predicting unknown graphemes. He then proceeds to a second stage where he apparently attempts to use recently acquired analytical processes for recognizing unknown printed words. But, if his analytical processes fail, he

does not use sentence context to predict the unknown word. Instead, he tends to give no response (Biemiller, 1970). Apparently, his recently acquired emphasis on analytical techniques for recognizing printed words leads him to exceed his memory capabilities for holding preceding ideas in mind. Hence, he cannot use contextual constraints. In short, for unknown words, he acts like a word-by-word reader.

With further progress and skill in word recognition processes, he proceeds to a third stage where he integrates his linguistic constraints and analytical processes for identifying printed words. Accomplishment of this integration is usually achieved by the second grade level (Biemiller, 1970), particularly if reading instruction starts out with connected discourse which capitalizes on natural language for word recognition (Barr, 1972). But, the natural language strategy, utilized in such reading programs as the language experience approach with its emphasis on whole-word recognition, must soon be supplemented with more analytic processes in order to increase effectiveness and efficiency in acquisition of word recognition ability (Samuels, 1970; Williams, 1970).

Thus, beginning readers learn to mobilize and integrate linguistic and perceptual processes in responding to print. These processes can be inferred from oral errors in responding to print. But, they can also be statistically determined by multiple regression analysis. Using this statistical procedure at the fourth grade level, phonological, semantic, morphological, and conceptual variables were found to account for some 90 percent of individual differences in attainment of power of reading. A perceptual-oculomotor variable also entered into prediction of speed of reading (Gilbert, 1953; Singer, 1965, 1969). At the sixth grade level, these predictors continued to account for individual differences in achievement of speed and power of reading. Affective factors, such as values (Athey, 1970, 1976; Athey and Holmes, 1969), also serve in the process of reading. They determine whether the individual is likely to mobilize the necessary systems for attempting to satisfy the demands of the reading task.

Although some abilities and processes continue to be involved in the structure and dynamics of reading, they cease to be first order, multiple-regression predictors of reading achievement when the group of readers involved in the statistical prediction samples have mastered them.

This interpretation explains why syntactic ability does not account for individual differences at the fourth grade level. This linguistic ability, with the exception of development of some sophisticated rules and versatility in grammatical control (C. Chomsky, 1970; Loban, 1965; Ruddell, 1970; Strickland, 1962), approaches maturity at age six when children usually begin formal reading instruction. In other words, if a group is *alike* in a certain function that underlies general reading ability, such as syntactic ability, that function does not account for individual differences in reading achievement at the first level of prediction. Or, when a group of readers becomes alike in a certain function, such as letter-name knowledge or letter-sound relationships, which are predictors of reading achievement in grade one (Murphy and Durrell, 1964), these functions, although they can still be mobilized in the reading process, do not remain as first-order predictors of general reading ability at subsequent grade levels.

Soon after the sixth grade level, word recognition and perceptual oculomotor control tend to drop out as first-level predictors. This is so because, at about this time, individuals generally tend to approach maturity in most word-recognition processes and in perceptual-oculomotor control (Gilbert, 1953). But, at the sixth grade, individuals tend to increase their vocabulary. This increase comes more from their reading than from their listening (Armstrong, 1953) because the vocabulary of literature has a greater range and depth than the vocabulary used for general, oral discourse. Hence, unlike word recognition processes, vocabulary ability continues to develop and remains as a predictor throughout the grades and at the college level (Holmes, 1954; Singer 1965).

Thus, the process of reading draws upon visual, perceptual, linguistic, and other cognitive processes plus certain values and personality components. But,

as individuals mature in the process of reading, individual differences in general reading ability shift from word recognition to semantic and morphological predictors.

As a variable, first-level predictors no longer account for individual differences in reading achievement, the relationship with IQ concomitantly diminishes for a particular age group. (Word recognition is a variable of this type.) But if a predictor becomes more variable for a particular group, then its relationship with IQ increases. (Vocabulary is a variable of this type.) Thus, the relationship between IQ and components of reading changes over the developmental span.

NATURE AND DEVELOPMENT OF IQ

IQ, operationally defined by the Stanford Binet test of intelligence, is the ratio of mental age to chronological age multiplied by 100. Mental age refers to the difficulty level of tasks that an individual can accomplish. Higher mental ages reflect ability to accomplish more verbally and more quantitatively abstract tasks. IQ has traditionally been defined as rate of learning or rate of past achievement (Smith and Dechant, 1961; Weir, 1967). Recently, however, IQ has been redefined as "developmental rate, the time required to arrive at a particular mental age." Developmental rate is then distinguished from "learning rate, the rate at which new information is acquired." Learning rate is a function of both MA (mental age) *and* IQ. That is, level of difficulty of tasks achieved and past rate of development are good predictors of tasks that can be attained and of rate of learning, provided conditions of instruction remain unchanged.

As used here, the *normal range of intelligence* is a continuum which excludes only organic retardates, those who have known organic defects (Zigler, 1967). Given adequate time for learning, such as relating letters to sounds, individuals within the normal range of intelligence can still master tasks which are within their mental age levels, albeit at variable rates. *In other words, if an individual is in the normal range of intelligence, he should be able to master the reading acquisition task, if it does not exceed his mental age and if he is given sufficient time for learning it.*

Like other facets of an individual's development, mental functions, as assessed by the Stanford Binet, also vary in time of *initial manifestation* and *rate of development* (Bayley, 1946, 1955). These latencies and asynchronies in the development of mental functions explain why the correlations of intelligence from preschool to adolescence are relatively low over the long time interval, but increase as the interval between ages of testing decrease.

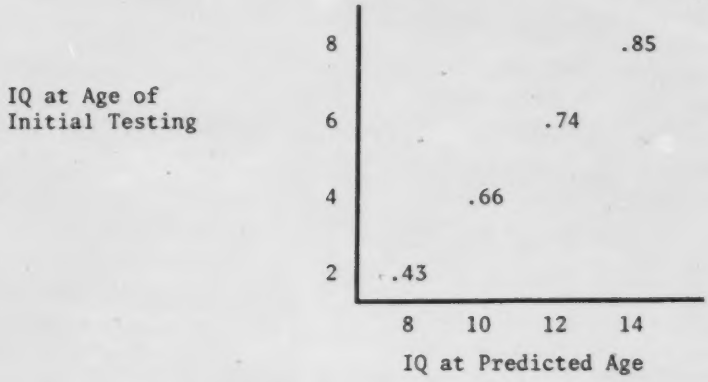
In general, the higher the age and the lower the interval between initial assessment and terminal assessment of IQ, the more likely the same functions are being correlated and hence the higher is the correlation. Since symbolic functions begin to be assessed about age four and since these functions increasingly enter into intelligence testing, correlations in IQ become increasingly higher and more stable at the older age intervals shown in Figure 2 on page 48, over a six-year range, for an age interval from two to eight, the correlation is .43, but from four to 10, it increases to .66, while from six to 12, it is even higher--.74, and from eight to 14 it is .85 (Honzik, MacFarlane, and Allen, 1948).

Changes in level of intellectual ability do occur. Over 50 percent of a group of individuals initially assessed at age six had changed by 15 or more IQ points and one-third of the group had changed over 20 IQ points by age 18 (Honzik et al, 1948). These magnitudes in IQ changes mean that individuals involved in this change would be reclassified over the 12-year span from "dull" to "average" or from "average" to "bright" or vice-versa.

Because of these changes in IQ, caution must be exercised in making predictions of intellectual performance, particularly over a long time interval. Consequently, use of intellectual ability as a criterion for reading expectancy

would also have to be modified. For example, the following principle for estimating expected level of reading achievement from a level of intellectual ability would have to be modified to take into account variation in performance due to socio-economic and ethnic status (as Harris [1972] did in stating the principle). However, the principle would also have to be modified to take into account the time gap between assessment of intelligence and measurement of reading achievement, as well as to take into account the criterion of intelligence used for estimating reading expectancy. The first part of the following principle is Harris'; the words in italics are mine: "A verbal intelligence scale, such as the Revised Stanford-Binet or the Wechsler Verbal I.Q., is still the best basis for estimating the level at which a child should be able to read with comprehension, *provided that the assessment of intelligence occurs about the same time as the estimate of the individual's reading level, and provided the criterion used for estimating expected reading achievement is mental age.*"

FIGURE 2



Correlations between IQ at Six-Year Intervals.
(after Honzik, MacFarlane and Allen, 1948)

Note that in this expectancy formula, Harris is careful to point out that he is estimating *reading with comprehension*, not just reading acquisition behavior. He is therefore including both reading acquisition behavior *and* general reading achievement in his estimation of expected level of reading achievement. If reading acquisition behavior alone were being estimated, then the IQ *and* the mental age criterion would have diminishing utility as expectancy criteria as members of a group progressed through school and learned to read.

Thus, throughout the normal range of intelligence, the correlation between IQ and reading or between mental age and reading should decrease as members of the group learn to read.

However, even though IQ may have some relation to reading acquisition behavior, it should be made explicit that IQ alone does not provide any useful information about reading achievement. For example, if you only knew that one individual had an IQ of 150 and another an IQ of 90, you would still not be able to make any valid statement about their reading ability. Only if you made an assumption about their chronological ages could you arrive at any valid estimate of their reading ability. In other words, it is necessary to know not only IQ but also chronological age so that mental age can be computed in estimating reading potential. Teachers usually have a particular age-group in mind when they correlate IQ with reading achievement. They are implicitly taking into account mental age and years of instruction in reading. Consequently, for predicting reading expectancy, teachers tend to operate both on IQ and MA, as advocated by Jensen and Rohwer (1968).

The implicit use of MA and years of schooling when correlating IQ with reading achievement is reflected throughout this paper in the use of the phrase "IQ for members of a particular age group" or its abbreviated version, "IQ for members of a group."

Although IQ may continue to be related to facilitation and application of learning, IQ for members of a group is not a general index of learning ability (Stake, 1961; Woodrow, 1949). Nor is it highly correlated with specific types of laboratory learning, such as verbal paired-associate learning--a type of learning that is involved in reading acquisition behavior.

IQ IN RELATION TO VERBAL LEARNING

Evidence on the relationship between IQ and learning has been summarized by Zeaman and House (1967). Paired-associate learning is related to school achievement, particularly when the paired-associate tasks involve verbal content (Duncanson, 1964; Rohwer, 1970) or abstract words (Stevenson, Hale, Klein, and Miller, 1968). The paired-associates paradigm has been successfully used in teaching some of the word recognition skills involved in the initial stages of learning to read (Gibson, 1965; Samuels, 1970; Singer, 1971; Williams, 1970).

Although verbal learning correlates well with verbal achievement, its relationship with IQ varies. The variability is to some degree attributable to conditions of instruction. Under some conditions, the relationship between IQ and various types of learning, including paired-associate learning, is low. Whether paired-associate learning is significantly related to IQ or not depends somewhat on the novelty of the task (McCullers, 1965). This type of stimulus task also affects the relation between IQ and paired-associate learning (Rohwer, 1970). In general, the correlations reported by Rohwer for IQ and paired-associate learning averaged only about .32.

Thus, under appropriate conditions of instruction, IQ can have a low relationship to verbal paired-associate learning. Since this type of learning can be used for teaching reading acquisition behavior, then by analogy individuals throughout the normal range of intelligence can learn to read. More direct evidence on the decreasing relationship between IQ and reading acquisition behavior can be gleaned from classroom studies of the relation between IQ and reading.

RELATION BETWEEN IQ AND READING

As we would expect, the relationship between IQ and reading achievement, as assessed by standardized, norm-referenced tests increases with grade in school. Durkin (1966) reported that the 49 children whose mean grade-equivalent reading achievement was 2.3 at the beginning of first grade ranged in Stanford Binet IQ from 91 to 161 with a median of 121. At the first grade level, IQ correlated only .40 with reading achievement. But in successive grades, the correlation increased and reached a magnitude of .79 at grade five.

An interpretation of this change from 16 to 64 percent common variance between IQ and reading achievement is the following: As reading achievement shifts from predominant emphasis on word recognition to stress on word meaning and comprehension, the mental functions being assessed by intelligence and reading tests have more in common. But, the low correlation at grade one also suggests that the one-to-one instructional conditions provided prior to grade one to the early readers in Durkin's study, plus other factors such as motivation to learn to read and parental expectation, may have diminished the relationship between IQ and reading achievement.

Expectancies also influence teacher behavior. In general, teachers tend to normalize instruction (Balow, 1964; Wilson, 1963). That is, they tend to modify instruction so that they can teach toward the average capability of the group. Hence, the stimulation given to learners and the amount learned is related to teacher expectation of student capability, especially when the content of the

curriculum is under teacher control. At least in the initial stages of reading, teachers are able to exercise some control over which words children are expected to learn.

To test the effect of teacher expectation on adaptation of the curriculum to the capabilities of the learner, Beez (1968) randomly assigned tutors and five- and six-year-old children to high- and low-expectancy conditions. The tutors were told their students were either expected to do well or poorly in school. This information then significantly influenced the number of words taught in the tutorial situation. Given 20 words to teach, the tutors in the high-expectancy condition actually taught an average of 10.4 words and their students learned to recognize 5.9 words, while the tutors in the low-expectancy condition taught on the average only 5.7 words and their students consequently learned to recognize only 3.7 words. There was no relationship between years of experience the tutors had in teaching and their performance in the tutorial situations. Nor was there any relation between student performance and their scores on the Peabody Picture Vocabulary Test. Thus, in this experiment, IQ was *not* related to number of words recognized.

Closely associated with reading achievement at the first grade level is *printed word perception*. That is, *achievement in a specific subject or process is a better predictor of future achievement in that subject than is achievement in general*. For example, Harrington and Durrell (1955) reported that visual and auditory perception of printed words had correlations of .64 and .56, respectively, with reading achievement at the end of grade one. On the other hand, Otis mental age had a correlation of only .23 with reading achievement at the end of grade one for Harrington and Durrell's subjects. Also, Gates (1926) discovered that after the Stanford Binet mental age had been partialled out from his sample of students in grades one to four, the correlation of *word perception* with reading was still as high as .69, which was higher than the zero-order correlation of .50 between intelligence and reading. Furthermore, his tests of associative learning correlated only about .33 with intelligence, and even lower with reading achievement. Gates concluded "...the [word] perceptive factor, irrespective of intelligence, is more closely associated with reading and spelling than all of the functions embraced in 'intelligence' as measured."

IQ for a particular group is more related to comprehension than to processes of reading. Using the matched mental age technique in which CA and IQ vary, Bliesmer (1952) compared 28 pairs of dull and bright children enrolled in public schools. The dull students were eighth and ninth graders whose Stanford Binet IQ's were 84 and below. The bright students were third and fourth graders whose IQ's were 116 and above. The mental age ranges were 10.7 to 12.6. The groups were not significantly different on reading rate, word recognition, perception of verbal relationship, reading for details, nor on word meaning subtests of the Durrell Analysis of Reading Difficulty. However, the bright students were significantly better on all tests of comprehension.

Dunn (1954) obtained similar results. He selected 20 mentally-retarded boys from special classes and matched them with 30 randomly selected boys from regular grades on Stanford Binet mental ages. The range in mental ages was 8.0 to 10.0. His results indicated that the regular group performed significantly better on silent and oral reading and use of context cues, but not on sound-blending ability or eye-movement behavior. Templin (1954) also found low relationships at the fourth grade level between Kuhlman-Anderson mental age and phonics tests: the correlations ranged from .29 for consonant sound discrimination to .43 for discrimination of sounds in words. Triggs, Cartee, Binks, Foster, and Adams (1954) reported that in grades four to six, WISC IQ correlations were .53 with comprehension, but only .15 with word recognition on the Diagnostic Survey Tests.

Thus, the relationship between IQ and reading acquisition for members of a group diminishes as individuals progress toward mastery of the process of learning to read.

METHODOLOGICAL CONSIDERATIONS IN RELATING IQ TO READING

The relationship between intellectual ability and reading also varies to some degree as a function of the measures used to assess them. Correlations between IQ and word recognition subtests could remain at a high level even in grades four, five, and six if the word recognition subtests require abstraction and generalization of symbol-sound relationships, or if the word recognition subtests are timed so that a premium is placed on speed of response. These conditions could account for the finding that correlation coefficients ranged from .34 to .72 between Stanford Binet mental age and Bond silent reading tests at the fourth grade level (Reynolds, 1953).

Subtests of intelligence also differentially vary in their relationships with reading subskills and general reading ability. Consequently, inability to read may not only hamper performance on reading tests, but also on the Stanford Binet intelligence test which contains items that are identical or almost identical to reading tasks. For example, Bond and Fay (1950) found that disabled readers at the fifth grade level are handicapped by five to 15 points because of inability to perform on the following Stanford Binet items: vocabulary, reading and remembering a story, abstract words, minkus completion, and dissected sentences.

Even more elements common to reading and intelligence occur when group tests of intelligence are used. Of course, the correlation between IQ and reading would be higher for these two types of tests. For example, Harootunian (1966) reported a correlation of .56 between California Achievement Tests and California Mental Maturity for seventh and eighth grades.

Other types of tests of intelligence vary considerably in their relationships with reading depending on the degree of commonality of tasks on the intelligence and the reading tests. Bond and Clymer (1955) reported that Primary Mental Abilities subtests of Space, Reasoning, Figure, Number, and Perception correlated only .08 to .43, while Verbal Words, Verbal Pictures, and Reasoning Words correlated .48 to .76 with Gates Test Type A (Reading to Appreciate General Significance) and D (Reading to Note Details).

In general, language functions assessed by intelligence tests are more closely related to reading than are non-language functions. Strang (1945) found that the non-language functions on the California Test of Mental Maturity correlated only .41 to .46 while the language functions correlated .80 to .84 with Thorndike-McCall and Gates Basic Reading Tests. Her interpretation of the results is that the two types of intelligence tests tap different mental functions and these mental functions are differentially related to reading achievement. Consistent with Strang's view, Belmont and Birch (1966) concluded from an analysis of use of language, particularly from definitions given to items on the WISC vocabulary subtest, that retarded readers were characterized by "inadequacy in language functioning rather than in perceptual or manipulative skills."

Similar findings have been reported by Reed (1970), who pointed out that deficits associated with reading failure and with degree of failure are a function of the expectancy criterion used to assess capability. Employing the WISC, Gates Diagnostic Survey Test, and Iowa Silent Reading Test, plus neuropsychological tests of Tactual Performance, Reitan Color Form, Progressive Figures, and Trait-Making to 233 fifth graders, Reed demonstrated that the "patterns of deficits between expectancy and performance levels are related to the method of measurement."

For example, Reed found that the difference between the good and poor readers on the neurophychological subtests were lower when the groups were matched on Verbal and Performance IQ scores than when they were matched on Full Scale IQ scores. Reed warned that the percentages of poor readers in the diagnostic categories of "modality deficiencies, cognitive deficits, aptitude weaknesses, and the relation of verbal to performance abilities will vary according to method of identifying the retarded reader."

He further speculated that "a child's potential for reading is probably much more closely related to methods and materials used for teaching than some arbitrary index of expectancy." Although Reed presented no evidence to support his speculation, his view is nevertheless consistent with Carroll's (1963) theory of schooling and Bloom's (1971) mastery learning strategies.

THEORY OF SCHOOL LEARNING AND MASTERY LEARNING STRATEGIES

Most of the studies reviewed thus far on the relationship between IQ and reading were based upon traditional age-graded schools where tasks to be learned and time for learning them are segmented into intervals of a year. Under these conditions, degree of attainment of a given task is measured after a fixed interval of time for all learners and correlated with predictors, including IQ tests, administered at the beginning of the time interval. Consequently, the correlations between IQ and specific subskills involved in the processes of reading acquisition are usually based upon inadequate time for some members of the group to learn them. Even so, the relationship between IQ and these reading acquisition subskills became increasingly lower as individuals learned how to read. But, they would have been even lower had some students in the group had more time to learn them. This is so because variability in achievement of the reading acquisition subskills within the group would then have decreased, and where variability decreases on either or both variables, the correlation between them must decrease. Thus, *time* is a significant component in learning to read, and in relating IQ to reading.

Time to learn is stressed in the elements of Carroll's (1963) model of school learning. Three of the elements are within the learner. They are "aptitude--the amount of time needed to learn a task under optimal conditions, ability to understand instruction [which is a function of his general intelligence and verbal ability], and perseverance--the time he is willing to engage actively in learning."

The other two elements are in the following conditions of instruction: opportunity--"time allowed for learning"--and quality of instruction. Quality of instruction is a function of the degree to which the tasks to be learned have been properly sequenced, presented, and adapted to the individual's "special needs and characteristics" and his "stage of learning." Time for learning in Carroll's model is inversely related to quality of instruction. Thus, the relationship between aptitude and achievement is a function not only of characteristics within the student, but also of conditions of instruction.

Applying Carroll's model of school learning to the curriculum, Bloom (1971) pointed out that if aptitude and achievement measures are reliable and valid, the expected correlation between them is about .70 or higher where the aptitude is normally distributed in the population and *the conditions of instruction are the same for all learners.* But, Bloom (1971) theorized that "the relation between aptitude and achievement should approach zero" if Carroll's model is applied to each student because about 95 percent of the population could then "learn a subject up to a high level of mastery." The five percent excluded would be those who had a special disability for learning a particular subject.

For example, Bloom refers to evidence that "selected criterion scores achieved by the top students at one grade level are achieved by the majority of students at another grade level." Also, where individuals learn at their own rates, they tend to achieve mastery at different time intervals. Bloom concluded that mastery learning is most appropriate where subjects are "*required, sequential,* and *closed,*" and which emphasize *convergent thinking.* He pointed out that such subjects important to the individual or society, require mastery type learning for cumulative progress, and can be mastered because they are characterized by a finite set of behaviors.

Mastery learning requires the use of criterion-referenced tests to assess progress. Under conditions of mastery learning and criterion-referenced tests,

IQ for members of a particular group would probably only be related to time to master the task because at least 95 percent of the group could eventually master the task. This type of learning and assessment contrasts sharply with fixed intervals of time for learning and with norm-referenced tests, respectively. If students' achievement is measured when they are in the middle stages of learning closed-objective tasks, then the relationship of achievement with IQ would be higher than when the tasks were mastered by the group. But, even norm-referenced tests for closed objectives, such as symbol-sound relationships, should have zero correlations with IQ, if achievement is assessed after students have been given varying amounts of time to master these closed objectives. For closed objectives, regardless of type of assessment, IQ should eventually be related only to rate of acquisition for members of a group (provided that students are given sufficient time to achieve the objective).

Reading acquisition behavior appears to be susceptible to mastery learning strategies and to criterion-referenced testing. Under this type of learning, the correlation between IQ and learning to read should change from some initially high level to zero for students throughout the range of normal intelligence. The literature on IQ and reading, reviewed above, provides some indirect evidence to support this view, but more direct evidence is needed.

The closest schools have come to the ideal of providing for individual differences in rate of reading acquisition behavior is through programmed instruction. Ellson, Harris, and Barber (1968) reported that the reading achievement of first grade children, given two daily 15-minute sessions of programmed tutoring, was significantly better than the control group which did not have the supplemental instruction. More significantly, the slow learners benefitted most; they almost equaled the range in achievement for the average students in the control group.

In a second study discussed by Ellson et al, programmed tutoring given twice daily produced significant improvement in reading achievement. Although children throughout the normal ability range improved, the gains were greatest for the low achievers. Moreover, Peabody Picture Vocabulary Test scores, which can be used as an estimate of intellectual ability, had a predictive coefficient of only .11 with the sight-word recognition test and .24 with word analysis, but .52 with comprehension. Thus, it appears that supplementing classroom instruction with programmed tutoring which provides a ratio of one teaching aide to one student, careful directions, systematic sequencing of stimuli, step-by-step assessment, individually-determined rate of progress, and variable time for learning is a defensible model for a mastery strategy in school learning and for further reduction in the relationship between IQ and learning to read.

Why Teachers Believe IQ Is Related to Learning To Read

If IQ and reading acquisition behavior for members of a group tend to have a decreasing relationship, and could have been a lower relationship, as students progress through the grades, why do teachers tend to believe that the relationship is higher? Some reasons can be generated from Carroll's model of school learning, from Bloom's concept of mastery learning strategies for the curriculum, and from ways in which IQ and reading behavior are defined and assessed. An additional reason could also be based on the "conventional wisdom" of maturational determinism for reading achievement that has prevailed over the past 30 to 40 years (Durkin, 1968). This maturational determinism was supported by Morphett and Washburne (1931), who adopted the maturationist position to explain their Winnetka results, and their study has had a widespread impact on teachers' instructional belief systems (Singer, 1970).

Despite conditions in the Winnetka school district, which ironically had a reading curriculum of 21 graded steps that today might be defined as mastery-type curriculum, Morphett and Washburne did not emphasize the variation in time children needed in learning to read. Instead, they computed that mental age cor-

related .50 with reading progress, and observed that children who had attained a mental age of six years, six months prior to beginning reading made more satisfactory progress than the "less mature children." That is, the successful children had progressed through 13 steps of the Winnetka graded program and had learned a minimum of 37 sight words before February of the first grade year. Generalizing their findings to *all* tests of intelligence, programs of instruction, and criteria for successful progress in reading, they advocated that "...by postponing the teaching of reading until children reach a mental age of six-and-a-half years, teachers can greatly decrease the chances of failure and discouragement, and can correspondingly increase their efficiency."

Since the mental age they advocated also tended to equal the average chronological age of the group, it was only one step more to divide chronological age by mental age and then relate an IQ of 100 or more to success in learning to read in the first grade.

In contrast, Gates and Bond (1936) invoked an environmentalist explanation for *their* results. They found that in some New York City schools (where teachers used a textbook and supplementary material for teaching four large classes of first grade students whose median IQ as 98.6), the correlation between Stanford Binet mental age and reading achievement at the end of the year was about .25. Those pupils who were making the *least* progress in reading made "marked" improvement after three weeks of special instruction while still in first grade. From these results, Gates and Bond concluded that determining the optimum time to begin instruction in reading was not a question of maturation, but a question of determining when the "maximum general and social returns" would accrue from learning to read at any given time. The following statement sums up their environmentalist position for reading readiness: "...the optimum time of beginning reading is not entirely dependent upon the nature of the child himself... it is in large measure determined by the nature of the reading program. We think there is no ultimate justification for assumming [sic] that materials and methods of teaching must remain forever fixed as they are, waiting upon nature to change the child through maturity until he reaches a point at which he can proceed successfully. We think, on the other hand, that techniques and materials of reading can be adjusted to teach children successfully at the time when reading is, all things considered, of optimum value to them." Thus, Gates and Bond sum up the conditions under which IQ is and is not related to beginning instruction.

SUMMARY AND CONCLUSION

Whether IQ is or is not related to reading for members of a particular age group depends on such variables as the nature and difficulty of the task, the capabilities of the reader, the time allowed for learning, the quality of instruction, and the nature of the tests used for assessing intelligence and reading.

The nature of the reading task can be broken down into its constituent components. At least five components are predictive of individual differences in reading achievement: (a) *Syntax* tends to approach complete development about age six. Graphophonological and other aspects of (b) *word recognition*, such as functional oculomotor efficiency, tend to reach a mastery level about ages 12 to 14. The other components of reading achievement, such as (c) *morphological processes*, (d) *word meaning processes*, and (e) *reasoning-in-reading processes*, continue to develop throughout a person's lifetime.

These five components can be accommodated under three major headings: word recognition, word meaning, and reasoning-in-reading. If the word meaning and reasoning-in-reading aspects of the reading task are kept within the mental age range of members of a group, then the relationship between IQ and reading acquisition behavior would tend to decrease toward zero as members of the group learned to read and to master the processes of word recognition in context. But if word

meaning and reasoning-in-reading continually increase in difficulty on the reading test so that they constantly challenge the most apt members of a group, then IQ will be highly related to reading achievement throughout acquisition, and even throughout the skilled stages of reading development.

The remedy is clear: separate assessment of reading acquisition from development of general reading ability. Criterion-referenced tests could be used to assess the former and norm-referenced tests to measure the latter.

Thus, the relationship between IQ for members in the normal intellectual range (as assessed by an individual measuring instrument, such as the Stanford Binet) and reading will at least be minimized if: (a) *the difficulty of the task is within the capabilities of the learner*; (b) *individual differences in rate are taken into account*; (c) *reading acquisition instruction is organized to facilitate sequential and cumulative learning*; (d) *directions are given clearly so that all students understand them*; and (e) *assessment of progress is based upon criterion-referenced tests.* On the other hand, if all of these variables are at the other extreme of their continua, and particularly if group tests of intelligence and norm-referenced tests of reading achievement are used, then the relationship between intelligence and reading achievement for members of a group will be maximized.

Thus, variations in the nature of the reading task, stage of reading development, conditions of instruction, and methods of assessment determine whether IQ for members of a group is or is not related to reading.

References

Aiello, J. and S. Jones. "Field Study of the Proxemic Behavior of Young Children in Three Subcultural Groups." *Journal of Personality and Social Psychology* 19:351-356, 1971.

Alland, A. *Human Diversity*. New York: Anchor Press, 1971.

Allitto, S. "The Language Issue in Communist Chinese Education." In *Aspects of Chinese Education*, edited by C. Hu. New York: Teachers College Press, 1969.

Armstrong, H.C. "The Relationship of the Auditory and Visual Vocabularies of Children." PhD dissertation [unpublished], Stanford University, 1953.

Athey, I. "Affective Factors in Reading." In *Theoretical Models and Processes of Reading*, edited by H. Singer and R. B. Ruddell, pp. 98-119. Newark, DE: International Reading Association, 1970.

_____. "Reading Research in the Affective Domain." In *Theoretical Models and Processes of Reading*, edited by H. Singer and R.B. Ruddell, pp. 352-380. Newark, DE: International Reading Association, 1976.

_____ and J.A. Holmes. *Reading Success and Personality Characteristics in Junior High School Students*. Berkeley, CA: University of California Press, 1969.

Baker, E. "The Effects of Manipulated Item Writing Constraints on the Homogeneity of Test Items." *Journal of Educational Measurement* 8:305-309, 1971.

Balow, I.H. "The Effects of Homogeneous Grouping in Seventh Grade Arithmetic." *The Arithmetic Teacher*, 1964, pp. 186-191.

Barr, R.C. "The Influence of Instructional Conditions on Word Recognition Errors." *Reading Research Quarterly* 7:509-529, 1972.

Bayley, N. "Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years." *Journal of Genetic Psychology* 75:165-196, 1949.

_____. "On the Growth of Intelligence." *The American Psychologist* 10:808-818, 1955.

Beez, W.V. "Influence of Biased Psychological Reports on Teacher Behavior and Pupil Performance." In *International Review of Research in Mental Retardation* (vol. 3), edited by N. Ellis. New York: Academic Press, 1968. [Summarized in "Educational Research in Mental Retardation," by S.L. Guskin and H.H. Specker.]

Belmont, L.C. and H.G. Birch. "The Intellectual Profile of Retarded Readers." *Perceptual Motor Skills* 22:787-876, 1966.

Bereiter, C. "An Academically Oriented Pre-school for Culturally Deprived Children." In *Pre-school Education Today*, edited by F. Hechinger. New York: Doubleday, 1966.

Biemiller, A. "The Development of the Use of Graphic and Contextual Information as Children Learn to Read." *Reading Research Quarterly* 6:75-96, 1970.

Bliesmer, E. "Reading Abilities of Bright and Dull Children of Comparable Mental Ages." *Journal of Educational Psychology* 45:321-331, 1952.

Bloom, B.S. "Mastery Learning and Its Implications for Curriculum Development." In *Confronting Curriculum Reform*,

edited by E.W. Eisner, chapter one. Boston: Little, Brown and Company, 1971.

Board, C. and D.R. Whitney. "The Effect of Selected Poor Item-writing Practices on Test Difficulty, Reliability and Validity." *Journal of Educational Measurement* 9(3):225-233, 1972.

Bond, G.L. and T.W. Clymer. "Interrelationship of the SRA Primary Mental Abilities, Other Mental Characteristics, and Reading Ability." *Journal of Educational Research* 49:131-136, 1955.

_____ and L.C. Fay. "A Comparison of the Performance of Good and Poor Readers on the Individual Items of the Stanford-Binet Scale, Forms L and M." *Journal of Educational Research* 43:475-479, 1950.

Brookover, W.B. and others. "Self Concept of Ability and School Achievement: II." *Improving Academic Achievement through Students' Self-concept Enhancement*. U.S. Office of Education, Cooperative Research Project No. 1636, 1965.

Byers, P. and H. Byers. "Nonverbal Communication and the Education of Children." In *Functions of Language in the Classroom*, edited by C. Cazden and others. New York: Teachers College Press, 1972.

Calfee, R.C. "Priorities for Reading Research in the Seventies." School of Education, Stanford University, 1974.

_____ and others. "How a Child Needs To Think To Learn To Read." In *Cognition in Learning and Memory*, edited by L. Gregg, pp. 139-182. New York: John Wiley and Sons, 1972.

_____ and R.L. Venezky. "Component Skills in Beginning Reading." In *Psycholinguistics and Reading*, edited by K.S. Goodman and J. Fleming, pp. 91-110. Newark, DE: International Reading Association, 1972.

Carroll, J.B. "A Model of School Learning." *Teachers College Record* 64:723-733, 1963.

Chomsky, C. "Reading, Writing, and Phonology." *Harvard Educational Review* 40:287-309, 1970.

Clark, K.B. *The Educationally Deprived*. New York: Metropolitan Applied Research Center, 1972.

Clark, M. *Reading Difficulties in Schools*. Baltimore: Penguin, 1970.

Cohen, R. "Conceptual Styles, Culture Conflict and Nonverbal Tests of Intelligence." *American Anthropologist* 71:828-856, 1969.

Cole, M. and J. Gay. "Culture and Memory." *American Anthropologist* 74:1066-1384, 1972.

Collier, J. *Alaskan Eskimo Education*. New York: Holt, Rinehart and Winston, 1973.

Cook, W.W. "The Functions of Measurement in the Facilitation of Learning." In *Educational Measurement*, edited by E.F. Lindquist. Washington, DC: American Council on Education, 1951.

Cronbach, L.J. "Comments on Mastery Learning and Its Implications for Curriculum Development." In *Confronting Curriculum Reform*, edited by E.W. Eisner. Boston: Little, Brown and Company, 1971.

Davis, F.B. "Fundamental Factors of Comprehension in Reading." *Psychometrika* 9:185-197, 1944.

_____, ed. *The Literature of Research in Reading with Emphasis on Models*. New Brunswick, NJ: Rutgers University, 1971.

Deutsch, M. "The Disadvantaged Child and the Learning Process." In *Education in Depressed Areas*, edited by A. Passow. New York: Teachers College Press, 1963.

DeVos, G. and H. Wagatsuma. *Japan's Invisible Race*. Berkeley, CA: University of California Press, 1966.

Doman, G. *How To Teach Your Baby To Read*. New York: Random House, 1963.

Donoghue, J. "An Eta Community in Japan." In *Modern Japanese Society*, edited by Y. and T. Ishida. Berkeley, CA: McCutchan, 1971.

Downing, J. and D.V. Thackray. *Reading Readiness*. London: University of London, 1971.

Dumont, R. "Learning English and How To Be Silent: Studies in Sioux and Cherokee Classrooms." In *Functions of Speech in the Classroom*, edited by C. Cazden and others. New York: Teachers College Press, 1972.

Duncanson, J.P. *Intelligence and the Ability To Learn*. Princeton, NJ: Educational Testing Service, 1964.

Dunn, L.M. "A Comparison of the Reading Processes of Mentally Retarded and Normal Boys of the Same Mental Age." *Child Development Publications* 19:7-99, 1954.

Durkin, D. *Children Who Read Early*. New York: Teachers College Press, 1966.

_____. "When Should Children Begin To Read?" In *Innovation and Change in Reading Instruction*, edited by H.M. Robinson. Chicago: University of Chicago Press, 1968.

_____. *Teaching Them To Read*. Boston: Allyn and Bacon, 1970.

_____. "Confusion and Misconceptions in the Controversy about Kindergarten Reading." In *The Formative Years: Principles of Early Childhood Education*, edited by S. Coopersmith and R. Feldman, pp. 228-235. San Francisco: Albion, 1974.

Educational Products Information Exchange Institute. *EPIE Education Product Report: Analysis of Basic and Supplementary Reading Materials. No. 64*. New York: Educational Products Information Exchange Institute, 1974.

Eidheim, H. "When Ethnic Identity Is a Social Stigma." In *Ethnic Groups and Boundaries*, edited by F. Barth. Boston: Little, Brown and Company, 1969.

Ellson, D.G. and others. "A Full Test of Programmed and Directed Tutoring." *Reading Research Quarterly* 3:307-367, 1968.

Farr, R. *Grade Levels and Test Scores*. Washington, DC: National Reading Center, 1971.

_____ and N. Anastasiow, eds. *Tests of Reading Readiness and Achievement: A Review and Evaluation*. Newark, DE: International Reading Association, 1969.

Fishman, J. and E. Lueders-Salmon. "What Has Sociology To Say to the Teacher?" In *Functions of Speech in the Classroom*, edited by C. Cazden and others. New York: Teachers College Press, 1972.

Gates, A.I. "A Study of the Role of Visual Perception, Intelligence, and Certain Associative Processes in Reading and Spelling." *Journal of Educational Psychology* 17:433-445, 1926.

_____ and G.L. Bond. "Reading Readiness: A Study of Factors Determining Success and Failure in Beginning Reading." *Teachers College Record* 37:678-685, 1936.

_____ and others. *Methods of Determining Reading Readiness*. New York: Teachers College Press, 1939.

Geyer, J.J. "Comprehensive and Partial Models Related to the Reading Process." *Reading Research Quarterly* 7:541-587, 1972.

Gibson, E. "Learning To Read." *Science* 148:1066-1072, 1965.

Gibson, E.J. and H. Levin. *The Psychology of Reading*. Cambridge, MA: The MIT Press, 1975.

Gilbert, L.C. "Functional Motor Efficiency of the Eyes and Its Relation to Reading." *University of California Publications in Education* 2:159-232, 1953.

Gladwin, T. *East Is a Big Bird*. Cambridge, MA: Harvard University Press, 1970.

Goffman, E. *Stigma*. Englewood Cliffs, NJ: Prentice-Hall, 1963.

Goodman, K.S. "A Linguistic Study of Cues and Miscues in Reading." *Elementary English*, 1965, 639-643.

Gumperz, J. "Sociolinguistics and Communication in Small Groups." *Language Research Laboratory Working Paper, No. 36.* Berkeley, CA: University of California, Berkeley, 1970.

_____ and E. Herasimchuk. "The Conversational Analysis of Social Meaning." In *Sociolinguistics: Current Trends and Prospects,* edited by R.W. Shuy. Washington, DC: Georgetown University Press, 1973.

Hall, E. "Environmental Communication." In *Behavior and Environment,* edited by A.E. Esser. New York: Plenum Press, 1971.

Harootunian, B. "Intellectual Abilities and Reading Achievement." *Elementary School Journal* 67:386-392, 1966.

Harrington, M.J. and D.D. Durrell. "Mental Maturity versus Perception Abilities in Primary Reading." *Journal of Educational Psychology* 46:375-380, 1955.

Harris, A.J. "Child Development and Reading." In *Readings on Reading Instruction,* edited by A.J. Harris and E.R. Sipay. New York: David McKay, 1972.

Henderson, E.H. and B.H. Long. "Personal-social Correlates of Academic Success among Disadvantaged School Beginners." *Journal of School Psychology* 9(2):101-113, 1971.

Hochberg, J. "Components of Literacy: Speculations and Exploratory Research." In *Basic Studies on Reading,* edited by H. Levin and J.P. Williams, pp. 74-89. New York: Basic Books, 1970.

_____ and V. Brooks. "Reading as an Intentional Behavior." In *Theoretical Models and Processes of Reading,* edited by H. Singer and R.B. Ruddell, pp. 304-314, 1970.

Holmes, J. "Factors Underlying Major Reading Disabilities at the College Level." *Genetic Psychology Monographs* 49:3-95, 1954.

Honzik, M.P. and others. "The Stability of Mental Test Performance between Two and Eighteen Years." *Journal of Experimental Education* 17:309-324, 1948.

Hoover, M.R. "Comments on Tests of Reading Readiness and of Initial Reading Skills." In *Final Report of the Conference on Tests of Reading Readiness and Initial Reading Skills,* edited by S.F. Wanat. Washington, DC: Georgetown University, 1974.

Hostetler, J. and G. Huntington. *Children in Amish Society.* New York: Holt, Rinehart and Winston, 1971.

Hughes, A.E. "An Investigation of Some Sociolinguistic Phenomena in the Vocabulary, Pronunciation and Grammar of Detroit Pre-School Children, Their Parents and Teachers." EdD dissertation, Michigan State University, 1967.

Jackson, P. *Life in the Classroom.* New York: Holt, Rinehart and Winston, 1968.

Jensen, A.R. and W.D. Rohwer, Jr. "Mental Retardation, Mental Age, and Learning Rate." *Journal of Educational Psychology,* 59:402-403, 1968.

Kirkland, M.C. "The Effects of Tests on Students and Schools." *Review of Educational Research* 41:303-350, 1971.

Kling, M. "Quest for Synthesis." In *The Literature of Research in Reading with Emphasis on Models,* edited by F.B. Davis, pp. 2-1 to 2-18. New Brunswick, NJ: Rutgers University, 1971.

Kolers, P.A. "Experiments in Reading." *Scientific American* 227:84-91, 1972.

Labov, W. "Phonological Correlates of Social Stratification." *American Anthropologist,* 1964A, 164-176.

_____ "Stages in the Acquisition of Standard English." In *Social Dialects and Language Learning,* edited by R. Shuy. Champaign, IL: National Council of Teachers of English, 1964B.

_____ "The Logic of Nonstandard English." *Florida FL Reporter* 7:60-75, 169, 1969.

_____ "Academic Ignorance and Black Intelligence." *The Atlantic* 229(6):59-67, 1972.

_____ and C. Robins. "A Note on the Relation of Reading Failure to Peergroup Status in Urban Ghettos." *Florida FL Reporter* 7:54-57, 167, 1969.

Landau, M. "Linkage, Coding and Intermediacy." *Journal of Comparative Administration* 3:401-429, 1971.

Lee, D. "Lineal and Nonlineal Codifications of Reality." *Psychosomatic Medicine*, 1950, 89-97.

Leonard, C. "A Method of Film Analysis of Ethnic Communication Style." Paper presented to the American Ethnological Society Meetings, Montreal, Canada, 1972.

Lewis, L. "Culture and Social Interaction in the Classroom." *Language-Behavior Research Laboratory Working Paper, No. 38*. Berkeley, CA: University of California, Berkeley, 1970.

Loban, W.D. *The Language of Elementary School Children*. Champaign, IL: National Council of Teachers of English, 1963.

Lowell, R.E. "A Factor Analysis of Reading Readiness Tests." *Journal of the New England Reading Association* 5(2):28-30, 1970.

MacGinitie, W.H. "Evaluating Readiness for Learning to Read: A Critical Review and Evaluation of Research." *Reading Research Quarterly*, 4(3):396-410, 1969.

Mackworth, J.F. "Comments on Test for Review." In *Final Report on the Conference on Tests of Reading Readiness and Initial Reading Skills*, edited by S.F. Wanat. Washington, DC: Georgetown University, 1974A.

_____ "What Early Reading Tests Should Measure." In *Final Report of the Conference on Tests of Reading Readiness and Initial Reading Skills*, edited by S.F. Wanat. Washington, DC: Georgetown University, 1974B.

Mackworth, N.H. "Comments on Test for Review." In *Final Report of the Conference on Tests of Reading Readiness and Initial Reading Skills*, edited by S.F. Wanat. Washington, DC: Georgetown University, 1974A.

_____ "What Early Reading Tests Should Measure." In *Final Report of the Conference on Tests of Reading Readiness and Initial Reading Skills*, edited by S.F. Wanat. Washington, DC: Georgetown University, 1974B.

Makita, K. "The Rarity of Reading Disability in Japanese Children." *American Journal of Orthopsychiatry* 38:599-614, 1968.

Marso, R.N. "Test Item Arrangement, Testing Time, and Performance." *Journal of Educational Measurement* 7:113-118, 1970.

McCullers, J.C. "Correlates of Verbal Paired Associate Learning in Children." *Psychological Reports* 17:747-752, 1965.

McDermott, R. "Achieving School Failure: An Anthropological Approach to Illiteracy and Social Stratification." In *Education and Cultural Process*, edited by G. Spindler. New York: Holt, Rinehart and Winston, 1974.

McMorris, R.F. and others. "Effects of Violating Item Construction Principles." *Journal of Educational Measurement* 9:287-295, 1972.

Mitchell, R. *The Korean Minority in Japan*. Berkeley, CA: University of California Press, 1967.

Morphett, M.V. and C. Washburne. "When Should Children Begin To Read?" *Elementary School Journal* 31 (March): 496-503, 1931.

Murphy, H.A. and D.D. Durrell. *Diagnostic Reading Readiness Test*. New York: Harcourt, Brace and World, 1964.

Nelson, J.B. "How the Reading Teacher Can Help Parents." In *The Quest for Competency in Teaching Reading*, edited by H.A. Klein, pp. 131-140. Newark, DE: International Reading Association, 1972.

Otto, W. "Evaluating Instruments for Assessing Needs and Growth in Reading." In *Assessment Problems in Reading*, edited by W.H. MacGinitie, pp. 14-20. Newark, DE: International Reading Association, 1973.

Palardy, J.M. "For Johnny's Reading Sake." *Reading Teacher* 22:720-724, 1969.

Payne, J. "Analysis of Teacher-Student Classroom Interaction in Amish and Non-Amish Schools." *Social Problems* 19:79-90, 1972.

Pribram, K.H. *Languages of the Brain*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Purkey, W.W. *Self Concept and School Achievement*. Englewood Cliffs, NJ: Prentice-Hall, 1970.

Reed, D. "Linguistic Forms and the Process of Reading." In *Basic Studies on Reading*, edited by H. Levin and J. Williams, pp. 19-29. New York: Basic Books, 1970.

Reynolds, M.C. "A Study of the Relationships between Auditory Characteristics and Specific Silent Reading Abilities." *Journal of Educational Research* 46:439-449, 1953.

Rist, R. "Student Social Class and Teacher Expectations." *Harvard Educational Review* 40:411-451, 1970.

Roberts, J. *Scene of the Battle: Group Behavior in the Classroom*. New York: Doubleday, 1970.

Rohwer, W.D. "Mental Elaboration and Proficient Learning." In *Minnesota Symposium on Child Learning* (vol. 4), edited by J.P. Hill. Minneapolis, MN: University of Minnesota Press, 1970.

_____. "Learning, Race, and School Success." *Review of Educational Research* 41:191-210, 1970.

Rubenstein, D. and D. Gajdusek. *A Study in Nascent Literacy*. Bethesda, MD: National Institute of Health, 1970.

Ruddell, R.B. "Psycholinguistic Implications for a Systems of Communication Model." In *Theoretical Models and Processes of Reading*, edited by H. Singer and R.B. Ruddell, pp. 239-258. Newark, DE: International Reading Association, 1970.

Rude, R.T. "Readiness Tests: Implications for Early Childhood Education." *The Reading Teacher* 26:572-580, 1973.

Samuels, S.J. "Modes of Word Recognition." In *Theoretical Models and Processes of Reading*, edited by H. Singer, and R.B. Ruddell, pp. 23-37. Newark, DE: International Reading Association, 1970.

Scheflen, A. "Living Space in an Urban Ghetto." *Family Process* 10:429-450, 1971.

_____. *Body Language and the Social Order*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Shimahara, N. *Burakumin: A Japanese Minority and Education*. The Hague: Martinus Nijhoff, 1971.

Shuy, R.W. "Some Language and the Cultural Differences in a Theory of Reading." In *Psycholinguistics and the Teaching of Reading*, edited by K. Goodman and J. Fleming. Newark, DE: International Reading Association, 1969.

_____ and others. *Sociolinguistic Factors in Speech Identification*. Research Project No. MH15048-01. Washington, DC: Center for Applied Linguistics, 1969.

Singer, H. *Substrata-Factor Reorganization Accompanying Development of General Reading Ability at the Elementary School Level*. Final Report, Contract No. 2011, U.S. Office of Education, 1965.

_____. "Theoretical Models of Reading." *Journal of Communication* 19:134-156, 1969.

_____. "Research that Should Have Made a Difference." *Elementary English* 47:27-34, 1970.

_____. "Theories, Models, and Strategies for Learning To Read." In *The Literature of Research in Reading with Emphasis on Models*, edited by F. Davis. New Brunswick, NJ: Rutgers University, 1971.

_____. "Translating Basic Resources into Classroom Practices." Paper presented at the Annual Meeting of the International Reading Association, Detroit, MI, 1972.

_____ and R.B. Ruddell, eds. *Theoretical Models and Processes of Reading*. Newark, DE: International Reading Association, 1970; 1976.

Smith, F. "Phonology and Orthography: Reading and Writing." *Elementary English* 49:1075-1088, 1972.

Smith, H.P. and E. Dechant. *Psychology in Teaching Reading*. Englewood Cliffs, NJ: Prentice-Hall, 1961.

Spindler, G. "The Transmission of American Culture." In *Education and Culture*, edited by G. Spindler. New York: Holt, Rinehart and Winston, 1963.

Stake, R.E. "Learning Parameters, Aptitudes, and Achievements." *Psychometric Monograph*, 1961, No. 9.

Stevenson, H.W. and others. "Interrelations and Correlates in Children's Learning and Problem Solving." *Monographs of the Society for Research in Child Development* 33(7):1-68, 1968.

Sticht, T.G. "Test Review." *In Final Report of the Conference on Tests of Reading Readiness and Initial Reading Skills*, edited by S.F. Wanat. Washington, DC: Georgetown University, 1974.

Strang, R. "Variability in Reading Scores on a Given Level of Intelligence Test Scores." *Journal of Educational Research* 38:440-446, 1945.

Strickland, R. "The Language of Elementary Textbooks and the Quality of Reading of Selected School Children." *Bulletin of the School of Education*, Indiana University 38:1-131, 1962.

Templin, M.C. "Phonic Knowledge and Its Relation to the Spelling and Reading Achievement of Fourth Grade Pupils." *Journal of Educational Research* 47:441-454, 1954.

Thompson, L. *Reading Disability*. Springfield, MA: Charles C. Thomas, 1966.

Thorndike, R.L. "Dilemmas in Diagnosis." In *Assessment Problems in Reading*, edited by W.H. MacGinitie, pp. 57-67. Newark, DE: International Reading Association, 1973.

Triggs, F.O. and others. "The Relationship between Specific Reading Skills and General Ability at the Elementary and Junior-Senior High School Levels." *Educational and Psychological Measurement* 14:176-185, 1954.

Valentine, C. "Deficit, Difference and Bicultural Models of Afro-American Behavior." *Harvard Educational Review* 41:137-158, 1971.

Wallace, A. "Driving to Work." In *Context and Meaning in Cultural Anthropology*, edited by M. Spiro. New York: Free Press, 1965.

Wanat, S.F. "Eye-fixation Patterning and Grammatical Structure." In *The Analysis of Reading Skill: A Program of Basic and Applied Research*, edited by H. Levin and others, pp. 378-379. Ithaca, NY: Cornell University, 1968.

_____. "Linguistic Structure in Reading." In *The Literature of Research in Reading with Emphasis on Models*," edited by F. Davis, pp. 8-155 to 8-177. New Brunswick, NJ: Rutgers University, 1971.

_____. "Effects of Language Patterns upon Scanning Patterns in Reading." In *Proceedings of the 80th American Psychological Association, 1972*, 7:533-534. Washington, DC: American Psychological Association, 1972A.

_____. "Performance Tests: Assessing Teachers of Reading." *Reading Teacher* 25:623-633, 1972B.

_____. "Developing Skills for Criticizing Research." *Reading Research Quarterly* 8, 1973.

_____. *Analysis of Reading Tests: Final Report*. Washington, DC: Georgetown University, 1974.

_____. "Reading Readiness Tests and ESOL." In *On TESOL*, edited by R. Crymes and W.E. Norris, pp. 201-205. Washington, DC: Teachers of English to Speakers of Other Languages, 1975.

_____. "Language behind the Eye: Some Findings, Speculations, and Research Strategies." In *Eye Movements and Psychological Processes*, edited by R.A. Monty and J.W. Senders, pp. 523-529. Hillsdale, NJ: Lawrence Erlbaum Associates, 1976A.

_____. "Relations between Language and Visual Processing." In *Theoretical Models and Processes of Reading*, edited

by H. Singer and R.B. Ruddell, pp. 108-136. Newark, DE: International Reading Association, 1976B.

_____ and H. Levin. "The Eye-Voice Span: Reading Efficiency and Syntactic Predictability." In *The Analysis of Reading Skill: A Program of Basic and Applied Research*, edited by H.levin and others, pp. 237-253. Ithaca, NY: Cornell University, 1968.

Weber, R.M. "A Linguistic Analysis of First Grade Reading Errors." *Reading Research Quarterly* 5:427-451, 1970.

Weintraub, S. "What Research Says to the Reading Teacher: Readiness Measures for Predicting Reading Achievement." *The Reading Teacher* 20:551-558, 1967.

Weir, M.W. "Mental Retardation, Technical Comment." *Science* 157:567, 1967.

Wiener, M. and W. Cromer. "Reading and Reading Difficulty: A Conceptual Analysis." *Harvard Educational Review* 47:620-643, 1967.

Williams, J.P. "Reaction to Modes of Word Recognition." In *Theoretical Models and Processes of Reading*, edited by H. Singer and R.B. Ruddell, pp. 38-46. Newark, DE: International Reading Association, 1970.

_____. "Six Papers in Search of a Model." In *The Literature of Research in Reading with Emphasis on Models*, edited by F.B. Davis, pp. 4-1 to 4-5. New Brunswick, NJ: Rutgers University, 1971.

Wilson, A. "Social Stratification and Academic Achievement." In *Education in Depressed Areas*, edited by A.H. Passow, pp. 217-235. New York: Teachers College Press, 1963.

Wolfram, W. "Sociolinguistic Implications for Educational Sequencing." In *Teaching Standard English in the Inner City*, edited by R. Fasold and R. Shuy. Washington, DC: Center for Applied Linguistics, 1970.

Woodrow, H.A. "The Ability To Learn." *Psychological Review* 53:147-158, 1946.

Woodson, M.I.C.E. "How Test Scales Get Their Meaning." School of Education, University of California, Berkeley, 1973.

Zeaman, D. and B.J. House. "The Relation of IQ and Learning." In *Learning and Individual Differences*, edited by R.M. Gagne, pp. 192-212. Columbus, OH: Charles E. Merrill, 1967.

Zigler, E. "Familial Mental Retardation: A Continuing Dilemma." *Science* 155:292-298, 1967.

Papers in Applied Linguistics

LINGUISTICS AND READING
SERIES: 1

1. Issues in Evaluating Reading
   edited by Stanley F. Wanat

2. Language and Reading Comprehension
   edited by Stanley F. Wanat

Roger W. Shuy, General Series Editor